SEMESTER I

Discipline: Computer Science and Engineering

Stream: CS2 (Artificial Intelligence and Data Science, Computational Linguistics, Data Science)

221TCS100	ADVANCED MACHINE LEARNING	CATEGORY	L	T	P	CREDIT
		DISCIPLINE CORE 1	3	0	0	3

Preamble: This course introduces machine learning concepts and popular machine learning algorithms. It will cover the standard and most popular supervised learning algorithms including linear regression, logistic regression, decision trees, k-nearest neighbour, an introduction to Bayesian learning and the naive Bayes algorithm, support vector machines and kernels and basic clustering algorithms. Dimensionality reduction methods and some applications to real world problems will also be discussed. It helps the learners to develop application machine learning based solutions for real world applications.

Course Outcomes:

After the completion of the course the student will be able to: *

CO 1	Analyse the Machine Learning concepts, classifications of Machine Learning algorithms and basic parameter estimation methods. **(Cognitive Knowledge Level: Analyse)**
CO 2	Illustrate the concepts of regression and classification techniques **(Cognitive Knowledge Level: Apply)**
CO 3	Describe unsupervised learning concepts and dimensionality reduction techniques. **(Cognitive Knowledge Level: Apply)**
CO 4	Explain Support Vector Machine concepts and graphical models. **(Cognitive Knowledge Level: Apply)**
CO 5	Choose suitable model parameters for different machine learning techniques and to evaluate a model performance. **(Cognitive Knowledge Level: Apply)**
CO6	Design, implement and analyse machine learning solution for a real-world problem. **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and developmentwork in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓		✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	
CO 5	✓		✓	✓	✓	✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	60-80%
Analyse	20-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation : 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/Quiz : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted.

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (Such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Suppose that X is a discrete random variable with the following probability mass function: where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations were taken from such a distribution: *(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)*. What is the maximum likelihood estimate of θ.

X	0	1	2	3
$P(X)$	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

2. What is the difference between Maximum Likelihood estimation (MLE) and Maximum a Posteriori (MAP) estimation?

3. A gamma distribution with parameters ***α, β*** has the following density function, where ***Γ(t)*** is the gamma function.

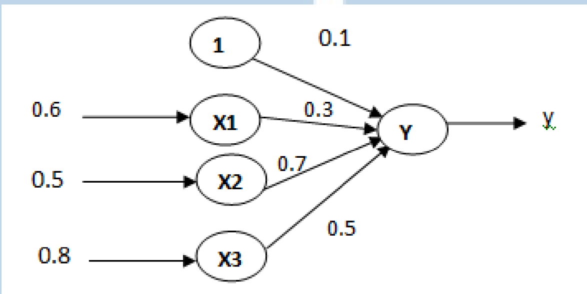
$$p(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

If the posterior distribution is in the same family as the prior distribution, then we say that the prior distribution is the conjugate prior for the likelihood function. Using the Gamma distribution as a prior, show that the Exponential distribution is a conjugate prior of the Gamma distribution. Also, find the maximum a posteriori estimator for the parameter of the Exponential distribution as a function of α and ***β***.

Course Outcome 2 (CO2)

1. How can we interpret the output of a two-class logistic regression classifier as a robability?

2. Calculate the output of the following neuron Y if the activation function is a binary sigmoid.



3. 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.

4. 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.

5. Why does a single perceptron cannot simulate simple XOR function? Explain how this limitation is overcome?

6. Consider a naive Bayes classifier with 3 boolean input variables, **X1**, **X2** and **X3**, and one boolean output, **Y**. How many parameters must be estimated to train such a naive Bayes classifier? How many parameters would have to be estimated to learn the above classifier if we do not make the naive Bayes conditional independence assumption?

Course Outcome 3(CO3):

1. Describe the basic operation of k-means clustering.

2. A Poisson distribution is used to model data that consists of non-negative integers. Suppose you observe m integers in your training set. Your model assumption is that each integer is sampled from one of two different Gaussian distributions. You would like to

learn this model using the EM algorithm. List all the parameters of the model. Derive the E-step and M-step for this model.

3. A uni-variate Gaussian distribution is used to model data that consists of non-negative integers. Suppose you observe m integers in your training set. Your model assumption is that each integer is sampled from one of two different Gaussian distributions. You would like to learn this model using the EM algorithm. List all the parameters of the model. Derive the E-step and M-step for the model.

4. Suppose you want to cluster the eight points shown below using **k**-means

	A_1	A_2
x_1	2	10
x_2	2	5
x_3	8	4
x_4	5	8
x_5	7	5
x_6	6	4
x_7	1	2
x_8	4	9

Assume that **k = 3** and that initially the points are assigned to clusters as follows:

C1 = {x1, x2, x3}, C2 = {x4, x5, x6}, C3 = {x7, x8}. Apply the **k**-means algorithm until convergence, using the Manhattan distance.

Course Outcome 4 (CO4):

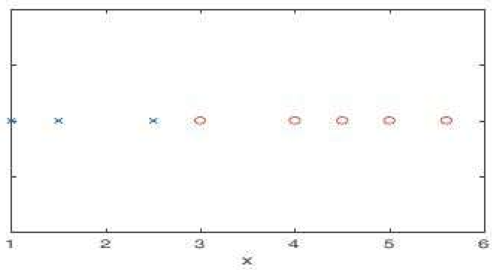
1. Describe how Support Vector Machines can be extended to make use of kernels. Illustrate with reference to the Gaussian kernel $K(x, y) = e^{-y}$, where $y = (x-y)^2$.

2. Suppose that you have a linear support vector machine (SVM) binary classifier. Consider a point that is currently classified correctly, and is far away from the decision boundary. If you remove the point from the training set, and re-train the classifier, will the decision boundary change or stay the same? Justify your answer.

3. What is the primary motivation for using the kernel trick in machine learning algorithms?

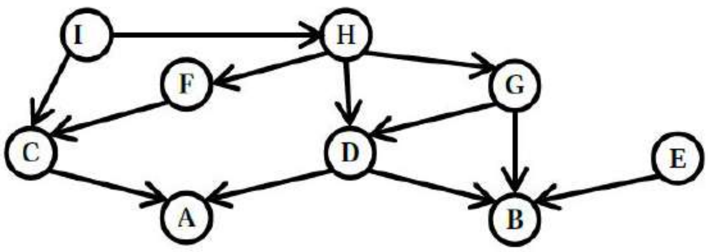
4. Show that the Boolean function $(x_1 \wedge x_2) \vee (\neg x_1 \wedge \neg x_2)$ is not linearly separable (i.e. there is no linear classifier **sign** $(w_1 x_1 + w_2 x_2 + b)$ that classifies all 4 possible input points correctly). Assume that "true" is represented by 1 and "false" is represented by −1. Show that there is a linear separator for this Boolean function when we use the kernel $K(x, y) = (x \cdot y)^2$ (*x.y* denotes the ordinary inner product) . Give the weights and the value of ***b*** for one such separator.

5. Consider the following one-dimensional training data set, ’x’ denotes negative examples and ’o’ positive examples. The exact data points and their labels are given in the table. Suppose a SVM is used to classify this data. Indicate which are the support vectors and mark the decision boundary. Give the value of the cost function and of the model parameters after training.



x	1	1.5	2.5	3	4	4.5	5	5.6
y	-1	-1	-1	1	1	1	1	1

6. Write down the factored conditional probability expression that corresponds to the graphical Bayesian Network shown below.



7. How do we learn the conditional probability tables(CPT) in Bayesian networks if information about some variables is missing? How are these variables called?

Course Outcome 5 (CO5):

1. Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the accuracy, precision and recall for the data.

2. Given the following data, construct the ROC curve of the data. Compute the AUC.

Threshold	TP	TN	FP	FN
1	0	25	0	29
2	7	25	0	22
3	18	24	1	11

4	26	20	5	3
5	29	11	14	0
6	29	0	25	0
7	29	0	25	0

3. With an example classification problem, explain the following terms: a) Hyper parameters b) Training set c) Validation sets d) Bias e) Variance.

4. What is ensemble learning? Can ensemble learning using linear classifiers learn classification of linearly non-separable sets?

5. Describe boosting. What is the relation between boosting and ensemble learning?

6. Classifier A attains 100% accuracy on the training set and 70% accuracy on the test set. Classifier B attains 70% accuracy on the training set and 75% accuracy on the test set. Which one is a better classifier. Justify your answer.

7. What are ROC space and ROC curve in machine learning? In ROC space, which points correspond to perfect prediction, always positive prediction and always negative prediction? Why?

8. Suppose there are three classifiers A, B and C. The (FPR, TPR) measures of the three classifiers are as follows – A (0, 1), B (1, 1) , C (1,0.5). Which can be considered as a perfect classifier? Justify your answer.

9. What does it mean for a classifier to have a high precision but low recall?

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES: 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M. TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221TCS100

Course Name: ADVANCED MACHINE LEARNING

Max. Marks: 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Explain the principle of the gradient descent algorithm.
2. In a two-class logistic regression model, the weight vector **w = [4, 3, 2, 1, 0]**. We apply it to some object that we would like to classify; the vectorized feature representation of this object is **x = [−2, 0, −3, 0.5, 3]**. What is the probability, according to the model, that this instance belongs to the positive class?
3. Expectation maximization (EM) is designed to find a maximum likelihood setting of the parameters of model when some of the data is missing. Does the algorithm converge? If so, do you obtain a locally or globally optimal set of parameters?
4. What is the basic idea of a Support Vector Machine?
5. What is the trade-off between bias and variance? **(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6. Suppose $x_1, \ldots, x_n$ are independent and identically distributed(iid) samples from a distribution with density **(7)**

$$f_X(x \mid \theta) = \begin{cases} \dfrac{\theta x^{\theta-1}}{3^\theta}, & 0 \le x \le 3 \\ 0, & \text{otherwise} \end{cases}$$

Find the maximum likelihood estimate (MLE) for θ.

7. Derive the gradient descent training rule assuming for the target function $o_d = w_0 + w_1x_1 + \ldots + w_nx_n$. Define explicitly the squared cost/error function E, assuming that a set of training examples D is provided, where each training example $d \in D$ is associated with the target output t_d. **(7)**

8. Cluster the following eight points representing locations into three clusters: ***A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)***. **(7)**

Initial cluster centers are: ***A1(2, 10), A4(5, 8)*** and ***A7(1, 2)***.

The distance function between two points ***a* = *(x1, y1)*** and ***b* = *(x2, y2)*** is defined as D***(a, b)*** = |***x2*** − ***x1***| + |***y2*** − ***y1***|

Use **k**-Means Algorithm to find the three cluster centers after the second iteration.

9. Describe Principal Component Analysis. What criterion does the method minimize? What is the objective of the method? Give a way to compute the solution from a matrix ***X*** encoding the features. **(7)**

10. Consider a support vector machine whose input space is 2-D, and the inner products are computed by means of the kernel ***K (x, y) = (x.y + 1)²-1*** (***x.y*** denotes the ordinary inner product). Show that the mapping to feature space that is implicitly defined by this kernel is the mapping to 5-D given by **(7)**

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow \phi(\mathbf{x}) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}\, x_1 x_2 \\ \sqrt{2}\, x_1 \\ \sqrt{2}\, x_2 \end{bmatrix}.$$

11. How does random forest classifier work? Why is a random forest better than a decision tree? **(7)**

12. 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. Compute the confusion matrix, accuracy, precision, recall, sensitivity and specificity on the following data. **(7)**

Sl.No.	**Actual**	**Predicted**
1	man	woman
2	man	man
3	woman	woman
4	man	man
5	man	woman
6	woman	woman
7	woman	man
8	man	man
9	man	woman
10	woman	woman

Syllabus

Module-1 (Parameter Estimation and Regression) 8 hours

Overview of machine learning: supervised, semi-supervised, unsupervised learning, reinforcement learning. Basics of parameter estimation: Maximum Likelihood Estimation (MLE), Maximum a Posteriori Estimation (MAP). Gradient Descent Algorithm, Batch Gradient Descent, Stochastic Gradient Descent. Regression algorithms: least squares linear regression, normal equations and closed form solution, Polynomial regression.

Module-2 (Regularization techniques and Classification algorithms) 9 hours

Overfitting, Regularization techniques - LASSO and RIDGE. Classification algorithms: linear and non-linear algorithms, Perceptrons, Logistic regression, Naive Bayes, Decision trees. Neural networks: Concept of Artificial neuron, Feed-Forward Neural Network, Back propagation algorithm.

Module-3 (Unsupervised learning) 8 hours

Unsupervised learning: clustering, k-means, Hierarchical clustering, Principal component analysis, Density-based spatial clustering of applications with noise (DBSCAN). Gaussian mixture models: Expectation Maximization (EM) algorithm for Gaussian mixture model.

Module-4 (Support Vector Machine and Graphical Models) 7 hours

Support vector machines and kernels: Max margin classification, Nonlinear SVM and the kernel trick, nonlinear decision boundaries, Kernel functions. Basics of graphical models - Bayesian networks, Hidden Markov model - Inference and estimation.

Module-5 (Evaluation Metrics and Sampling Methods) 8 hours

Classification Performance Evaluation Metrics: Accuracy, Precision, Precision, Recall, Specificity, False Positive Rate (FPR), F1 Score, Receiver Operator Characteristic (ROC) Curve, AUC. Regression Performance Evaluation Metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R Squared/Coefficient of Determination. Clustering Performance Evaluation Metrics: Purity, Jaccard index, Normalized Mutual Information, Clustering Accuracy, Silhouette Coefficient, Dunn's Index. Boosting: AdaBoost, gradient boosting machines. Resampling methods: cross-validation, bootstrap. Ensemble methods: bagging, boosting, random forests Practical aspects in machine learning: data preprocessing, overfitting, accuracy estimation, parameter and model selection Bias-Variance tradeoff

Course Plan

No	Topics	No. of Lectures (40)
1	**Module-1 (Parameter Estimation and Regression) 8 hours**	
1.1	Overview of machine learning: supervised, semi-supervised, unsupervised learning, reinforcement learning.	1
1.2	Basics of parameter estimation: Maximum Likelihood Estimation(MLE)	1
1.3	Basics of parameter estimation: Maximum Likelihood Estimation(MLE) - Examples	1
1.4	Basics of parameter estimation: Maximum a Posteriori Estimation (MAP)	1
1.5	Basics of parameter estimation: Maximum a Posteriori Estimation (MAP) - Example	1
1.6	Gradient Descent Algorithm, Batch Gradient Descent, Stochastic Gradient Descent	1
1.7	Regression algorithms: least squares linear regression, normal equations and closed form solution	1
1.8	Polynomial regression	1
2	**Module-2 (Regularization techniques and Classification algorithms) 9 hours**	

2.1	Overfitting, Regularization techniques - LASSO and RIDGE	
2.2	Classification algorithms: linear and non-linear algorithms	
2.3	Perceptrons	
2.4	Logistic regression	
2.5	Naive Bayes	
2.6	Decision trees	
2.7	Neural networks: Concept of Artificial neuron	
2.8	Feed-Forward Neural Network	
2.9	Back propagation algorithm	
3	**Module-3 (Unsupervised learning) 8 hours**	
3.1	Unsupervised learning: clustering, k-means	
3.2	Hierarchical clustering	
3.3	Principal component analysis	
3.4	Density-based spatial clustering of applications with noise (DBSCAN)	
3.5	Gaussian mixture models: Expectation Maximization (EM) algorithm for Gaussian mixture model	
3.6	Gaussian mixture models: Expectation Maximization (EM) algorithm for Gaussian mixture model	
4	**Module-4 (Support Vector Machine and Graphical Models) 7 hours**	
4.1	Support vector machines and kernels: Max margin classification	
4.2	Support vector machines: Max margin classification	
4.3	Nonlinear SVM and the kernel trick, nonlinear decision boundaries	
4.3	Kernel functions	
4.5	Basics of graphical models - Bayesian networks	
4.6	Hidden Markov model - Inference and estimation	
4.7	Hidden Markov model - Inference and estimation	
4.8	Hidden Markov model - Inference and estimation	
5	**Module-5 (Evaluation Metrics and Sampling Methods) 8 hours**	
5.1	Classification Performance Evaluation Metrics: Accuracy, Precision, Precision, Recall, Specificity, False Positive Rate (FPR), F1 Score, Receiver Operator Characteristic (ROC) Curve, AUC	
5.2	Regression Performance Evaluation Metrics: Mean Absolute Error	

	(MAE), Root Mean Squared Error (RMSE), R Squared/Coefficient of Determination	
5.3	Clustering Performance Evaluation Metrics: Purity, Jaccard index, Normalized Mutual Information, Clustering Accuracy, Silhouette Coefficient, Dunn's Index	
5.4	Boosting: AdaBoost, gradient boosting machines.	
5.5	Resampling methods: cross-validation, bootstrap.	
5.6	Ensemble methods: bagging, boosting, random forests	
5.7	Practical aspects in machine learning: data preprocessing, overfitting, accuracy estimation, parameter and model selection	
5.8	Bias-Variance tradeoff	

Reference Books

1. Christopher Bishop. Neural Networks for Pattern Recognition, Oxford University Press, 1995.

2. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective, MIT Press 2012.

3. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements Of Statistical Learning, Second edition Springer 2007.

4. Ethem Alpaydin, Introduction to Machine Learning, 2nd edition, MIT Press 2010.

5. Tom Mitch

CODE 221TCS003	MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE	CATEGORY	L	T	P	CREDIT
		Program Core 1	3	0	0	3

Preamble:

This course is intended to provide the learners with an outlook on applying concepts in linear algebra in the fields of data science, machine learning, and artificial intelligence. This course helps the learners to acquire the skills to implement the concepts in MATLAB/Python and then apply linear algebra concepts to real datasets. Also, this course discusses the Challenges of applying the acquired knowledge in different Optimization and Linear Algebra concepts toward the inference and prediction stages of Data Analytics.

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

CO 1	Analyse the fundamentals of linear algebra and calculus, and other mathematical concepts for Artificial Intelligence, Machine Learning, and Data Science (**Cognitive knowledge level: Analyse**)
CO 2	Apply the knowledge acquired in different optimization and linear algebra concepts towards the inference and prediction stages of data analytics. (**Cognitive knowledge level: Apply)**
CO 3	Implement linear algebra concepts in scientific programming languages (MATLAB, Python) **(Cognitive knowledge level: Apply)**
CO 4	Apply eigenvectors and SVD for Dimensionality reduction **(Cognitive knowledge level: Apply)**
CO 5	Design, Develop, implement, and Present innovative Ideas on the application of linear algebra for Data Science, Machine learning, and Artificial Intelligence **(Cognitive Knowledge Level: create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively and write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate mastery over the area as per the program's specialization. The mastery should be at a level higher than the requirements in the appropriate bachelor's program

PO4: An ability to apply stream knowledge to design or develop solutions for real-world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tools to model, analyze and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream-related problems taking into consideration sustainability, societal, ethical, and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✔	✔	✔	✔	✔	✔	
CO 2	✔	✔	✔	✔	✔	✔	
CO 3	✔	✔	✔	✔	✔	✔	
CO 4	✔	✔	✔	✔	✔	✔	
CO 5	✔	✔	✔	✔	✔	✔	✔

Assessment Pattern

Bloom's Category	End Semester Examination
Understand	
Apply	60-80%
Analyze	20-40%
Evaluate	Can be done through Assignments
Create	Can be done through Assignments

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

The evaluation shall only be based on application, analysis or design-based questions (for both internal and end-semester examinations).

Continuous Internal Evaluation : 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/Quiz : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects are not permitted.

The test paper shall include a minimum of 80% of the syllabus.

Course-based task/test paper questions shall be useful in testing the knowledge, skills, comprehension, application, analysis, synthesis, evaluation, and understanding of the students.

End Semester Examination Pattern:

The end-semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (Such questions shall be useful in testing knowledge, skills, comprehension, application, analysis, synthesis, evaluation, and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem-solving and quantitative evaluation), with a minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

The total duration of the examination will be 150 minutes.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Discuss Linear algebra handles large amounts of data,

2, List ten Powerful Applications of Linear Algebra in Data Science

3. 'Support Vector Machine is an application of the concept of Vector Spaces in Linear Algebra'- Investigate

4. Support Vector Machine Classification

Course Outcome 2 (CO2)

1. Write an algorithm for simple linear regression by gradient descent method

2. Implement Gradient Descent for multilinear regression from scratch

3. Elaborate regularized linear regression for model prediction and reduce errors

Course Outcome 3(CO3):

1. Elaborate on the different ways to compute and conceptualize matrix multiplication with examples

2. Implement Vector Hadamard multiplication

3. w1 = [1 3 5];

4. w2 = [3 4 2];in MATLAB/ Python

Course Outcome 4 (CO4):

1. Implement the determinant of a matrix product in Python

2. Elucidate the usefulness of Linear Algebra in Dimensionality Reduction

    i. Principal Component Analysis (PCA)

    ii. Singular Value Decomposition (SVD)

3. Determine whether the following matrices have a null space. If so, provide basis vector(s) for that null space.

$$a \begin{pmatrix} 4 & 3 \\ 1 & 1 \\ 0 & 5 \end{pmatrix} \qquad b. \begin{pmatrix} 3 & 1 & 5 \\ 4 & 1 & 0 \end{pmatrix}$$

Course Outcome 5 (CO5):

1. List the five steps of model fitting

2. Express the average or the mean of a set of numbers as a model and use least squares to fit that model

3. Elaborate on the usage of Linear Algebra in Machine Learning

    a. Loss functions

    b. Regularization

    c. Covariance Matrix

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M. TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221TCS003

Course Name: Mathematical Foundation for Data Science

Max. Marks: 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Explain normal equation. Contrast normal equation and gradient descent. **5**

2. Explain overfitting and the method of resolving overfitting **5**

3. Make use of the parallel matrix multiplication method to find AB, where **5**

$$A = \begin{bmatrix} 2 & 1 & 5 & 3 \\ 0 & 7 & 1 & 6 \\ 9 & 2 & 4 & 4 \\ 3 & 6 & 7 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 6 & 1 & 2 & 3 \\ 4 & 5 & 6 & 5 \\ 1 & 9 & 8 & -8 \\ 4 & 0 & -8 & 5 \end{bmatrix}$$

4. Let v=⟨1,3⟩ and w=⟨−4, −2⟩. Write v as the sum of two orthogonal vectors, one of which is the projection of v⟶ onto w **5**

5. Find the SVD for the matrix **5**

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$

(5x5=25)

Part B

(Answer any five questions. Each question carries 7 marks)

6. (a) Write an algorithm for simple linear regression by gradient descent method **(7)**

7. (a) Implement Gradient Descent for multilinear regression from scratch **(7)**

8. (a) Elaborate the different types of matrices with examples and implement them using MATLAB or NumPy **(7)**

i. Square matrix
ii. Rectangular matrix
iii. Symmetrix matrix
iv. Skey- symmetric matrices
v. Identity matrix
vi. Zero
vii. Diagonal matrix
viii. Triangular matrix
ix. Augmented
x. Complex

9. (a) Define rank and a maximum possible rank **(2)**

(b) Create a matrix of 10 X 10 with a rank of 4 (use matrix multiplication) **(3)**

(c) Generalize the procedure to create any M x N matrix with a rank r (2) **(2)**

10. (a) Implement the determinant of a matrix product in Python **(3)**

(b) Determine whether the following matrices have a null space. If so, provide basis vector(s) for that null space. **(4)**

a $\begin{pmatrix} 4 & 3 \\ 1 & 1 \\ 0 & 5 \end{pmatrix}$ $\quad$ *b.* $\begin{pmatrix} 3 & 1 & 5 \\ 4 & 1 & 0 \end{pmatrix}$

11. (a) Write MATLAB or Python code to implement the following experiment: **(7)**

a. Generate a 2 × 3 matrix of random numbers.
b. Compute its SVD.
c. Compute two eigen decompositions using the matrix and its transpose.
d. Confirm that the two sets of eigenvalues match, and check whether the eigenvalues match the singular values

Plot the eigenvectors and singular vectors in 2D or 3D (as appropriate) to confirm that SVD and transpose + eigen decomposition produce the same eigenspaces

12. (a) Determine the eigenvalues and the corresponding eigenvectors of the following matrix A = $\begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$ **(7)**

Syllabus (Emphasis shall be on Problem-solving, implementing concepts, and application- Languages Python/MATLAB)

Module	Contents	hours
I	**GRADIENT DESCENT AND REGULARIZATION** Gradient Descent, Intuitions, Gradient Descent for a Regression algorithm, Multiple Features, Gradient Descent on Multiple Features, Practice on Gradient Descent for Polynomial Regression, Normal Equation	8
II	**LINEAR ALGEBRA-VECTORS:** Vectors: geometry and algebra, Vector addition, and subtraction, Vector-scalar multiplication, Dot product geometry, Vector orthogonality, Cauchy-Schwarz inequality, Vector Hadamard multiplication, cross product, unit vectors, **VECTOR SPACES:** Dimensions, and fields in linear algebra, Subspaces, Subspaces vs. subsets, Span, Linear independence, Basis **MATRICES: introduction:** dimensionality, Matrix operations, Matrix-scalar multiplication, Implementation, Transpose, Complex matrices, Addition, equality, transpose, Diagonal, and trace,	7
III	**MATRICES-: MATRIX MULTIPLICATION-**Introduction, matrix multiplication by layering, Multiplication with diagonals, Matrix-vector multiplication, Symmetric matrices, multiply symmetric matrices Hadamard Multiplication, asymmetry Index, Code challenge, **RANK-**concepts, Maximum possible rank, Computing rank, Rank and scalar multiplication, Rank of added and multiplied matrices, Rank of A & A^T, A^T A, AA^T, random matrices, Boosting rank by shifting, rank difficulties, rank, and span, Code challenge: **MATRIX SPACES:** Column space and Row space of a matrix (A & AA^T), Null space of a matrix, orthogonal subspaces, Dimensions of column/row/null spaces, Example of the four subspaces, code challenge	7
IV	**DETERMINANTS, PROJECTIONS & ORTHOGONALIZATION-** **DETERMINANT-**Determinant, Determinant of a 2x2 matrix, Determinant of a 3x3 matrix, characteristic polynomial, the full procedure, determinant of triangles, determinant and row reduction, determinant and scalar multiplication, theory vs practice, Code challenge **MATRIX INVERSE:** Concept and applications, Inverse of a Diagonal matrix, Inverse of a 2x2 matrix, The MCA algorithm to compute the inverse, Computing the inverse via row reduction, Left inverse and right inverse, Pseudo-inverse, Code challenge **PROJECTIONS, AND ORTHOGONALIZATIONS:** Projections in R^2, Projections in R^N, Orthogonal and parallel vector components, Orthogonal matrices, Gram-Schmidt procedure, QR decomposition, Inverse via QR Decomposition, Code challenge **LEAST SQUARES FOR MODEL-FITTING IN STATISTICS**: Introduction, Least squares via left inverse, Least squares via orthogonal projection, Least-squares via row-reduction, Model-predicted values, and residuals, Least-squares applications, Code challenge	10
V	**DIMENSIONALITY REDUCTION: EIGEN DECOMPOSITION-** Eigenvalues, eigenvectors, Eigen decomposition, Diagonalization, Matrix powers via diagonalization, Distinct and repeated eigenvalues, symmetric	8

	matrices, Eigen layers of a matrix, Eigen decomposition of singular matrices, Matrix powers and Inverse, Generalized eigen decomposition, Code challenges **SINGULAR VALUE DECOMPOSITION (SVD):** Singular value decomposition, Computing the SVD, singular values and eigenvalues, Symmetric Matrices, SVD and the four subspaces, SVD, and matrix rank, Spectral theory of matrices, SVD for low-rank approximations, Normalizing singular values, the Condition number of a matrix, SVD and Matrix Inverse, MP pseudo inverse, code challenges	

Course Plan

S.NO	TOPIC	NO. OF LECTURES
MODULE 1 - GRADIENT DESCENT AND REGULARIZATION-8 hours		
1.1	Gradient Descent, Intuitions,	1
1.2	Gradient Descent for a Regression algorithm.	1
1.3	Gradient Descent for a Regression algorithm.	1
1.4	Multiple Features	1
1.5	Gradient Descent on Multiple Features,	1
1.6	Practice on Gradient Descent	1
1.7	Gradient Descent for Polynomial Regression	1
1.8	Gradient Descent and Normal Equation	1
MODULE 2 LINEAR ALGEBRA- 7 HOURS		
	LINEAR ALGEBRA- Vectors	1
2.1	**VECTORS:** geometry and algebra, Vector addition, and subtraction, Vector-scalar multiplication,	1
2.2	Dot product geometry, Vector orthogonality,	1
2.3	Vector Hadamard multiplication, cross product, unit vectors,	
2.4	**VECTOR SPACES:** Dimensions, and fields in linear algebra, Subspaces, Subspaces vs. subsets	1
2.5	Span, Linear independence, Basis	1
2.6	**MATRICES**: dimensionality, Matrix operations, Matrix-scalar multiplication,	1
2.7	Complex matrices, Addition, equality, transpose, Diagonal, and trace,	1
MODULE 3 -MATRICES – 7 HOURS		
3.1	**MATRIX MULTIPLICATION:** matrix multiplication by layering,	1

	Multiplication with diagonals, Matrix-vector multiplication,	
3.2	Symmetric matrices, multiply symmetric matrices Hadamard Multiplication, asymmetry Index, Code challenge,	1
3.3	**RANK-**concepts, Maximum possible rank, Computing rank, Rank and scalar multiplication, Rank of added and multiplied matrices	1
3.4	The rank of A & $A^{T,}$ A^T A, $AA^{T,}$ random matrices,	1
3.5	boosting rank by shifting, rank difficulties, rank, and span, Code challenge:	1
3.6	**MATRIX SPACES:** Column space and Row space of a matrix (A & AA^T), Null space of a matrix,	1
3.7	orthogonal subspaces, Dimensions of column/row/null spaces, Example of the four subspaces, code challenge,	1
MODULE 4- DETERMINANTS, & PROJECTIONS, AND ORTHOGONALIZATION- 10 HOURS		
4.1	**DETERMINANT-** Determinant of a 2x2 matrix, & 3x3 matrix, characteristic polynomial, the full procedure,	1
4.2	determinant of triangles, determinant and row reduction, determinant and scalar multiplication, theory vs practice, Code challenge	1
4.3	**MATRIX INVERSE:** Concept and applications, Inverse of a Diagonal matrix, Inverse of a 2x2 matrix, The MCA algorithm to compute the inverse,	1
4.4	Computing the inverse via row, Left inverse and right inverse, Pseudo-inverse, Code challenge	1
PROJECTIONS AND ORTHOGONALIZATION:		
4.5	Projections in R^2, Projections in R^N, Orthogonal and parallel vector components,	1
4.6	Orthogonal matrices, Gram-Schmidt procedure, QR decomposition,	1
4.7	Inverse via QR Decomposition, Code challenge	1
LEAST SQUARES FOR MODEL-FITTING IN STATISTICS		
4.8	Introduction, least squares via left inverse, Least squares via orthogonal projection,	1
4.9	Least squares via row-reduction, Model-predicted values, and residuals	1
4.10	Least-squares applications, Code challenge	1
MODULE 5- DIMENSIONALITY REDUCTION- 8 HOURS		
EIGEN DECOMPOSITION-,		
5.1	Eigenvalues, eigenvectors, Eigen decomposition,	1
5.2	Diagonalization, Matrix powers via diagonalization	1
5.3	Distinct and repeated eigenvalues, symmetric matrices, Eigen layers of a matrix,	1

5.4	Eigen decomposition of singular matrices, Matrix powers and Inverse, Generalized eigen decomposition, Code challenges	1
SINGULAR VALUE DECOMPOSITION (SVD) -		
5.5	Singular value decomposition, Computing the SVD, singular values and eigenvalues,	1
5.6	Symmetric Matrices, SVD and the four subspaces, SVD, and matrix rank,	1
5.7	Spectral theory of matrices, SVD for low-rank approximations, Normalizing singular values, the Condition number of a matrix,	1
5.8	SVD and Matrix Inverse, MP pseudo inverse, code challenges	1

TEXT BOOKS:

1. Mike X Cohen, Linear Algebra: Theory, Intuition, Code [Print Replica] Kindle Edition
2. Gene H. Golub, Charles F. Van Loan, “Matrix Computations”, John Hopkins University Press.
3. Steven Cooper, Data Science from Scratch: The #1 Data Science Guide for Everything A Data Scientist Needs to Know: Python, Linear Algebra, Statistics, Coding, Applications, Neural Networks, and Decision Tree Kindle Edition
4. Randolf H. Reiss, B.S, “Eigen Values and Eigen Vectors in Data Dimension Reduction for Regression”, San Marcos, Texas.
5. Gilbert Strang, “Linear Algebra and its Applications”, Thomson Learning Inc.

REFERENCES:

1. Charu C. Aggarwal, “Linear Algebra and Optimization for Machine Learning”, Springer 2020.
2. Singiresu S. Rao, “Engineering Optimization: Theory and Practice”, Fourth Edition 2009 by John Wiley & Sons, Inc.

221TCS004	INTRODUCTION TO AI AND NLP	CATEGORY	L	T	P	CREDIT
		Program Core 2	3	0	0	3

Preamble:

This course introduces the concepts, tools, and techniques of machine learning for text data. The students will learn the elementary concepts as well as emerging trends in the field of NLP. This course helps the learners to extract information from unstructured text, identify linguistic structure of it, and to apply different the techniques for text analytics. The learners will able to implement and evaluate NLP applications using machine learning and deep learning methods.

Course Outcomes:

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

CO1	Analyze the applications of AI in the domain of NLP **(Cognitive Knowledge Level: Apply)**
CO2	Transform text into an appropriate data structure (**Cognitive Knowledge Level: Apply)**
CO3	Apply Probability models, language models, and Markov models for Text processing **(Cognitive Knowledge Level: Apply)**
CO4	Build NLP applications using Machine Learning Methods **(Cognitive Knowledge Level: Apply)**
CO5	Design, Develop, Implement and Present innovative Ideas on NLP and AI **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real-world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tools to model, analyze and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓		✓	✓	
CO 2	✓		✓		✓	✓	
CO 3	✓		✓		✓	✓	
CO 4	✓	✓	✓	✓	✓	✓	
CO 5	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	50-80%
Analyse	20-40%

Evaluate	Assess using Assignments/Project
Create	Assess using Assignments/Project

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

The evaluation shall only be based on application, analysis or design-based questions (for both internal and end-semester examinations).

Continuous Internal Evaluation: 40 marks
Micro project/Course based project : 20 marks
Course based task/Seminar/Quiz : 10 marks
Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects are not permitted. Test paper shall include a minimum of 80% of the syllabus.

Course-based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation, and understanding of the students.

End Semester Examination Pattern:

The end-semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (Such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation, and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem-solving and quantitative evaluation), with a minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

The total duration of the examination will be 150 minutes.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Implement artificial intelligence-based solution in agriculture for optimization of irrigation and application of pesticides and herbicides

2. Develop Artificial Intelligence/Machine Learning based for Diabetes Care

3. Explain the role of Natural Language Processing Applications in Finance

Course Outcome 2 (CO2)

1. Which of the following techniques can be used for keyword normalization in NLP, the process of converting a keyword into its base form?

a. Lemmatization

b. Soundex

c. Cosine Similarity

d. N-grams

Interpret your answer.

2. Explain any two out of the eight methods that **Convert Text to Features**

a. Recipe 1. One Hot encoding.

b. Recipe 2. Count vectorizer.

c. Recipe 3. N-grams.

d. Recipe 4. Co-occurrence matrix.

e. Recipe 5. Hash vectorizer.

f. Recipe 6. Term Frequency-Inverse Document Frequency (TF-IDF)

g. Recipe 7. Word embedding.

h. Recipe 8. Implementing fast Text.

3. Explain the following with an example

a. Regular expression and its working

b. Properties of regular expression

c. Meta characters Big brackets [] and search [abcd] in ‘kasdfaiabcasdfaabc’ and write the output

d. Let’s look at the following sentence: “I ate an apple and played the piano.” Generate the one-hot embedding matrix for this sentence

Course Outcome 3(CO3):

a. Implement TF-IDF from scratch
b. Explain how you will implement NLP in other languages
c. Explain Markov Model for text classifier and build a text classifier using the Markov model

Example: Take two sets of poems by two different authors, Edgar Allan Poe and Robert Frost. Given a sentence, the system should be able to predict the author

Course Outcome 4 (CO4):

1. Application: Latent Semantic Indexing for Search Engine Optimization using PCA/SVD

2. Implement ANN for multiclass classification

3. Implement Spam Detection using Naïve Bayes or Logistic regression

Course Outcome 5 (CO5):

1. Implement Text Summarization using python

2. How will you implement a sentiment analyzer in Python using logistic regression to predict sentiment on Amazon reviews?

3. Application: Topic Modeling Using Latent Dirichlet Allocation

Model Question Paper

QP CODE:

Reg No: _______________

Name: __________________ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221TCS004

Course Name: INTRODUCTION TO AI AND NLP

Max. Marks: 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. a. Contrast NLTK and Spacy **5**

b. Explain the different types of artificial Intelligence

2. a. Analyse the importance of the concept of vectors in NLP **5**

b. Explain the simple functioning of the ELIZA chatbot

3. Explain Markov Property and Markov model for NLP **5**

4. Discuss topic modeling and the intuition behind using LDA for topic modeling **5**

5. Describe implementing binary text classification using TensorFlow **5**

(5x5=25)

Part B

(Answer any five questions. Each question carries 7 marks)

6. (a) Explain the applications of Artificial intelligence in Health care **(7)**

7. (a) Elaborate vectors, their usefulness in NLP, tokenization, stop words, stemming, and lemmatization with examples **(3)**

(b) Explain the limitations of the count vectorizer and the necessity of TF-IDF **(2)**

(c) In a corpus of N documents, one randomly chosen document contains a total of T terms, and the term “hello” appears K times. What is the correct value for the product of TF (term frequency) and IDF (inverse-document-frequency), if the term “hello” appears in approximately one-third of the total documents? **(2)**

8. (a) Explain the regular expression and need for a regular expression in NLP **(3)**

(b) How do regular expressions work? **(2)**

(c) Explain the use of common Regex functions used in NLP **(2)**

9. (a) Explain the language model, applications of language modeling, method to Compute the probability of a sentence, and the curse of dimensionality **(2)**

(b) what is Markov’s assumption in language modeling and n-grams **(3)**

(c) Implement n-grams and update-function **(2)**

10. (a) Explain the spam detection problem and why you want to detect them (2) **(3)**

(b) Apply the Markov model n-gram approach to solve this problem and implement in Python **(4)**

11 (a) Explain RNN for text classification in TensorFlow **(2)**

(b) Explain Parts of Speech (PoS)Tagging in TensorFlow **(3)**

(c) Explain Named Entity Recognition in TensorFlow **(2)**

12. (a) Explain CNN for text classification **(7)**

Syllabus

Module	Contents	Total Lecture Hours (40 hrs)
1	**INTRODUCTION TO ARTIFICIAL INTELLIGENCE:** Artificial Intelligence? History, AI on a conceptual level, Types of AI, Use Cases, importance and applications of AI, AI algorithms, types of machine learning, types of problems solved in AI, advantages, and disadvantages of AI, AI In Marketing, Banking, Finance, Agriculture, HealthCare, Gaming, Space Exploration., Autonomous Vehicles, Chatbots, Artificial Creativity, AI Tools & Frameworks, AI vs Machine Learning vs Deep Learning, an overview of python for AI, **INTRODUCTION TO NLP:** NLP in the Real World, NLP Tasks, Language? Its Building Blocks, Why Is NLP Challenging? Machine Learning, Deep Learning, and NLP: An Overview, Approaches to NLP, Heuristics-Based NLP, Machine Learning & Deep Learning for NLP, NLP Pipeline, Applications of NLP-Machine translation, Speech recognition, Image Captioning, spam detection, text prediction- Introduction to Software Packages-Spacy, NLTK, Gensim, PyTorch, **Regular Expression-** importance, properties, working and python package (re), case study: working of Eliza chatbot	7
2	**REGULAR EXPRESSION & TEXT PROCESSING:** Common regex function, Meta Characters- Big brackets, cap, Backslash, Squared Brackets, Special Sequences, Asterisk, Plus, And Question mark, Curly Brackets **Understanding Pattern Objects-** Match Method Vs Search Method, Finditer Method, Logical Or, Beginning And End Patterns, Parenthesis **String Modification-** split method, sub-method, subn method, **Text Processing**-Words, Tokens, Counting words, vocabulary, corpus, tokenization in spacy- **Sentiment Classification-** (yelp) download a review dataset use data preparation using NumPy, pandas, counter, re-add tokens to vocabulary, build vocabulary from a data frame, from corpus, one hot encoding, encoding documents, train test splits, feature computation, confusion matrix, analysis. **Language Independent Tokenization**: Types of tokenization — Word, Character, and sub-word tokenization, problems with word tokenizer, drawbacks of a character-based tokenizer, problems with sub-word tokenization, Byte Pair Encoding, , **String Matching and Spelling Correction**-Minimum edit distance- table filling, dynamic programming,	9
	WORD EMBEDDING & PROBABILISTIC MODELS: Vector Models & Text Preprocessing: Vectors, Bag of Words, Count Vectorizer, Tokenization, Stopwords, Stemming, and Lemmatization, Stemming, and Lemmatization, Count Vectorizer, Vector Similarity. TF-IDF, Word-to-Index Mapping, Building TF-IDF, Neural Word Embeddings, Neural Word Embeddings. Vector Models Text Pre-processing Summary, steps of NLP analysis, **Probabilistic Models-Language**	9

3	**Modelling:** importance, types of language modeling, the curse of dimensionality, Language Model Markov Assumption And N-Grams, Language Model Implementation – Setup, Ngrams Function, Update Counts Function, Probability Model Function, Reading Corpus, Language Model Implementation Sampling Text, **Markov Models:** Markov Property, Markov Model, Probability Smoothing and Log-Probabilities, Building a Text Classifier, **Article Spinning** – Problem, N-Gram Approach, implementation, **Cipher Decryption with Language Modeling And Genetic Algorithm** Ciphers, substitution cipher, bigrams, maximum likelihood, and log-likelihood, Language models, Genetic Algorithms,	
4	**NLP USING MACHINE LEARNING MODELS**-Spam Detection– Problem, Naive Bayes theorem, Intuition, spam detection using Naïve Bayes, class imbalance, ROC, AUC, AND F1 SCORE, Implementing spam detection in python, **Sentiment Analysis** -Problem, Logistic Regression Intuition, Multiclass Logistic Regression, Logistic Regression Training and Interpretation, sentiment analysis implementation in python, **Text Summarization**-Using Vectors, Text Rank Intuition,Text Rank in Python, Text Summarization in Python **Topic Modeling**-different topic modeling techniques, Latent Dirichlet Allocation (LDA) – Essentials, Latent Dirichlet Allocation–Topic Modeling with Latent Dirichlet, **Latent Symmatc Modelling(Indexing)**-LSA / LSI Introduction, Singular Value Decomposition Intuition, LSA / LSI: Applying SVD to NLP, Latent Semantic Analysis / Latent Semantic Indexing in Python	7
5	**DEEP LEARNING-** word embeddings, nonlinear neural networks, **Neuron** – Intro, Fitting a Line, Classification Code Preparation, Text Classification in Tensorflow, The Neuron, How does a model learn?, **Feed Forward Neural Networks**- Ann-introduction, The Geometrical Picture, Activation Functions, Multiclass Classification, Text Classification ANN in Tensorflow, Text Preprocessing Code Preparation, Text Preprocessing in Tensorflow, Embeddings, CBOW(continuous bag of words), CBOW in Tensorflow, **Convolution Neural Networks**- Convolution, pattern matching, weight sharing, convolution in color images, CNN Architecture, CNN for Text, CNN for NLP in Tensorflow, **Recurrent Neural Networks**- Simple RNN / Elman Unit, RNNs: Paying Attention to Shapes, GRU, and LSTM. RNN for Text Classification in TensorFlow, Parts-of-Speech Tagging, and Named Entity Recognition in TensorFlow	8

COURSE PLAN		
Sl.NO	**TOPIC**	**NO. OF LECTURES**
	Module 1 – Introduction of AI & NLP (7 hours)	
Module 1: Introduction To Artificial Intelligence (7 hours)		
1.1	Artificial Intelligence? History, AI on a conceptual level, Types of AI, Use Cases, importance and applications of AI,	**1**
1.2	AI algorithms, types of machine learning, types of problems solved in AI, advantages, and disadvantages of AI	**1**
1.3	AI In Marketing, Banking, Finance, Agriculture, HealthCare, Gaming, Space Exploration, Autonomous Vehicles, Chatbots,	**1**
1.4	Artificial Creativity, AI Tools & Frameworks, AI vs Machine Learning vs Deep Learning, an overview of python for AI,	**1**
	INTRODUCTION TO NLP,	
1.5	NLP in the Real World, NLP Tasks, Language? Its Building Blocks, Why Is NLP Challenging? Machine Learning, Deep Learning, and NLP: An Overview, Approaches to NLP, Heuristics-Based NLP, Machine Learning & Deep Learning for NLP	**1**
1.6	NLP Pipeline, Applications of NLP-Machine translation, Speech recognition, Image Captioning, spam detection, text prediction-	**1**
1.7	Introduction to Software Packages-Spacy, NLTK, Gensim, PyTorch, **Regular Expression-** importance, properties, working and python package (re), case study: working of Eliza chatbot	**1**
Module 2- Regular Expression & Text Processing (9 Hours)		
2.1	Common regex function used in NLP, -	**1**
2.2	Meta Characters- Big brackets, cap, Backslash,	**1**
2.3	Squared Brackets, Special Sequences, Asterisk, Plus, And Question mark, Curly Brackets	**1**
2.4	**Understanding Pattern Objects-** Match Method Vs Search Method, Finditer Method, Logical Or, Beginning And End Patterns, Parenthesis	**1**
2.5	**String Modification**- split method, sub-method, subn method,	**1**
	TEXT PROCESSING	
2.6	Words, Tokens, Counting words, vocabulary, corpus, tokenization in spacy-	**1**
2.7	**Sentiment Classification-** (yelp) download a review dataset use –data preparation using NumPy, pandas, counter, re-add tokens to vocabulary, build vocabulary from a data frame, from corpus, one hot encoding, encoding documents, train test splits, feature computation, confusion matrix, analysis.	**1**

2.8	**Language Independent Tokenization**: Types of tokenization — Word, Character, and sub-word tokenization, problems with word tokenizer, drawbacks of a character-based tokenizer, problems with sub-word tokenization, Byte Pair Encoding	**1**
2.9	**String Matching and Spelling Correction**-Minimum edit distance- table filling, dynamic programming,	**1**
Module 3-Word Embedding & Probabilistic Models (9 Hours)		
3.1	**Vector Models & Text Preprocessing:** Vectors, Bag of Words, Count Vectorizer, Tokenization, Stop words,	**1**
3.2	Stemming and Lemmatization, Count Vectorizer, Vector Similarity. TF-IDF,	**1**
3.3	Word-to-Index Mapping, Building TF-IDF	**1**
3.4	Neural Word Embeddings, Neural Word Embeddings Demo. Vector Models & Text Pre-processing Summary, steps of a typical NLP analysis	**1**
	PROBABLISTIC MODELS	
3.5	**Language Modelling:** types of language modeling, the importance of language modeling, the curse of dimensionality, Language Model Markov Assumption And N-Grams,	**1**
3.6	Language Model Implementation – Setup, Ngrams Function, Update Counts Function, Probability Model Function, Reading Corpus, Language Model Implementation Sampling Text,	**1**
3.7	**Markov Models:** Markov Property, Markov Model, Probability Smoothing and Log-Probabilities, Building a Text Classifier,	**1**
3.8	**Article Spinning**–Problem Description, N-Gram Approach, implementation in python,	**1**
3.9	**Cipher Decryption with Language Modeling And Genetic Algorithm-** Ciphers, substitution cipher, bigrams, maximum likelihood, and log-likelihood, Language models, Genetic Algorithms,	**1**
Module 4 – NLP Using Machine Learning Models (7 Hours)		
4.1	**Spam Detection**– Problem, Naive Bayes theorem, Intuition, spam detection using Naïve Bayes, class imbalance, ROC, AUC, AND F1 SCORE, Implementing spam detection in python,	**1**
4.2	**Sentiment Analysis** -Problem, Logistic Regression Intuition, Multiclass Logistic Regression,	**1**
4.3	Logistic Regression Training and Interpretation, sentiment analysis implementation in python	**1**
4.4	**Text Summarization**-Using Vectors, Text Rank Intuition	**1**
4.5	Text Rank in Python, Text Summarization in Python	**1**
4.6	**Topic Modeling**-different topic modeling techniques, Latent Dirichlet Allocation (LDA) – Essentials, Latent Dirichlet Allocation–Topic Modeling with Latent Dirichlet	**1**
4.7	**Latent Symmatc Modelling(Indexing)**-LSA / LSI Introduction, Singular Value Decomposition Intuition, LSA / LSI: Applying SVD to NLP, Latent Semantic Analysis / Latent Semantic Indexing in Python	**1**

Module 5 - Deep Learning (8 Hours)		
5.1	word embeddings, nonlinear neural networks	**1**
5.2	**Neuron** – Intro, Fitting a Line, Classification Code Preparation, Text Classification in Tensorflow, The Neuron, How does a model learn?,	**1**
5.3	**Feed Forward Neural Networks-** Ann-introduction, The Geometrical Picture, Activation Functions, Multiclass Classification, Text Classification ANN in Tensorflow,	**1**
5.4	Text Preprocessing Code Preparation, Text Preprocessing in Tensorflow, Embeddings, CBOW(continuous bag of words), CBOW in Tensorflow	**1**
5.5	CONVOLUTION NEURAL NETWORKS:- CNN-Introduction, Convolution, pattern matching, weight sharing, convolution in color images,	**1**
5.6	**Convolution Neural Networks-** Convolution, pattern matching, weight sharing, convolution in color images, CNN Architecture, CNN for Text, CNN for NLP in Tensorflow,	**1**
5.7	**Recurrent Neural Networks-** Simple RNN / Elman Unit, RNNs: Paying Attention to Shapes, , GRU, and LSTM. RNN for Text Classification in TensorFlow,	**1**
5.8	Parts-of-Speech Tagging, and Named Entity Recognition in TensorFlow	**1**

Reference Books

1. Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana, Practical Natural Language Processing, Shroff/O'Reilly, 2020.

2. Steven Bird, Ewan Klein, Edward Loper, Natural Language Processing with Python, O'Reilly

3. Akshay Kulkarni, Adarsha Shivananda, Natural Language Processing Recipes Unlocking Text Data with Machine Learning and Deep Learning using Python, Apress

4. Taweh Beysolow II 'Applied Natural Language Processing with Python- Implementing Machine Learning and Deep Learning Algorithms for Natural Language Processing, Apress

5. Palash Goyal, Sumit Pandey, Karan Jain 'Deep Learning for Natural Language Processing Creating Neural Networks with Python, Apress

6. Hobson Lane. Cole Howard, Hannes Max Hapke 'Natural Language Processing in Action, Understanding, analyzing, and generating text with Python', ©2019 by Manning Publications Co

8. Wolfgang Ertel, Introduction to Artificial Intelligence, Springer,

221ECS012	DATA ANALYTICS	CATEGORY	L	T	P	CREDIT
		PEC-2	3	0	0	3

Preamble:

This course enables the students to understand the concepts of Data Analytics. It covers Data and Relations, Correlation, Basic Data Analytics and visualization methods using R, Finite State Machines, Dimensionality reductions, Feature extraction, Clustering, Classification and Regression Techniques, and scalability through parallelization. It helps the learners to develop applications for real time data analysis.

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

CO 1	Identify data errors and dependencies among attributes by modelling them as sets & relations. **(Cognitive Knowledge Level: Apply)**
CO 2	Apply statistical methods for evaluation hypothesis (**Cognitive Knowledge Level: Apply)**
CO 3	Apply regression, classification, and clustering models on a given dataset **(Cognitive Knowledge Level: Apply)**
CO4	Apply correlation techniques to find the dependencies between the features. **(Cognitive Knowledge Level: Apply)**
CO5	Develop applications that uses the concepts in Data Analytics **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams.

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards.

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects.

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PO 7**
CO 1	✓		✓	✓	✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	
CO5	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	60-80%
Analyse	20-40%

Evaluate	
Create	

Assignments or course projects can be used for higher level assessment of course outcomes.

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks
iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. How to model stochastic and deterministic errors. Explain with examples.

2. What are the ways in which various errors can be handled?

Course Outcome 2 (CO2):

1. What R commands would you use to remove null values from a data set.

2. Which function R can be used to a fit nonlinear line to the data.

Course Outcome 3 (CO3):

1. Consider the data sets for two classes X1 = {(0,0)} and X2 = {(1,0), (0,1)}. Which classification probabilities will a naive Bayes classifier produce for the feature vector (0,0)?

2. Explain SVM classifier with an example.

Course Outcome 4 (CO4):

1. For the data set X = { (1,0), (2,0), (3,1) , (4,1), (5,1), (6,1), (7,0), (8,0) } compute chi-square test statistic for 4 bins.

2. Explain the difference between correlation and causality.

Course Outcome 5 (CO5):

1.Develop a small application for a manufacturing industry to maintain their works using the concepts in data analytics.

2. Develop a small application for improving Transportation System using the concepts in data analytics.

Model Question Paper

QP CODE:

Reg No: ______________

Name:_________________ **PAGES : 2**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M. TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS012

Course Name: DATA ANALYTICS

Max. Marks: 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Compute the output of (a) an asymmetric moving mean filter, q =3, (b) an asymmetric moving median filter, q = 3, (c) an exponential filter, y0 = 0, correction term n= 0.5 for the time series (0,0,0,1,0,0,0). Which filter result do you like best?
2. Differentiate between Type I and Type II errors.
3. Construct the data tuples for an autoregressive forecasting model with a time horizon of m = 2 for the time series x = (1,2,3,5,8).
4. Explain prototype-based clustering.
5. Explain Rectified linear activation unit. **(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	Explain stochastic and deterministic errors with examples. Using 2-sigma rule and m-sigma rule how a value is classified as outliers?	(7)
7.	Consider the data sets for two classes X1 = {(0,0)} and X2 = {(1,0), (0,1)}. Which classification probabilities will a naive Bayes classifier produce for the feature vector (0,0)?	(7)
8.	Illustrate the importance of visualizing data before analysis.	(7)
9.	What is feature scaling? Explain the feature scaling techniques.	(7)
10	Justify how the computational complexity of the nearest neighbor is reduced by the LVQ approach.	(7)
11	For the data set X = { (1,0), (2,0), (3,1), (4,1), (5,1), (6,1), (7,0), (8,0) } compute chi-square test statistic for 4 bins.	(7)
12	Consider the two-dimensional patterns (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8). Compute the principal component using PCA Algorithm. Use PCA Algorithm to transform the pattern (2, 1) onto the Eigen vector.	(7)

Syllabus: Error Handling, Correlation, Models, Clustering, Data and Process Parallelization, Batch processing frameworks.

	Syllabus	
Module	**Content**	**Hours**
1	**Data and Relations** - Data scales, Set and Matrix representations, Relations, Similarity and dissimilarity measures, Sequence relations. Data pre-processing - Error types, error handling, filtering, transformation, merging. **Correlation** - Linear, Causality, Chi-Square tests. Cross validation and feature selection.	**7**
2	**Basic Data Analytics Methods Using R -** Descriptive Statistics, Statistical methods for evaluation, Hypothesis Testing, ANOVA. **Visualization methods using R - Exploratory** Data Analysis, visualizing single Variable, Examining Multiple Variables	**9**
3	**Models-** Finite state machines, Recurrent models, Autoregressive models, Moving Average Models.	**6**
4	**Clustering** - Cluster partitions, Sequential clustering, Prototype based clustering, Relational clustering, Cluster tendency assessment, Cluster validity, Self-organising map (SOP). Use Cases **Regression-** Linear Regression, Logistic regression, Use Cases	**7**
5	**Classification Methods-** Naive Bayes classifier, Decision Trees, LDA, SVM, Learning Vector Quantization. **Scalability through parallelization** - Data parallelization, Process parallelization, Scaling using feature engineering, Dimensionality Reduction, Cascading, Feature reduction through spatial transforms. Global Average Pooling, Data Augmentation, **Case Studies:** ReLU nonlinearity, MLP, Convolutional Layer.	**11**

Course Plan

No	Topic	No. of Lectures
1	**Data and Relations, Correlation**	
1.1	Data scales, Set and Matrix representations	1
1.2	Relations, Similarity and dissimilarity measures	1
1.3	Sequence relations.	1
1.4	Data pre-processing - Error types, error handling, Filtering	1
1.5	Transformation, merging	1
1.6	Correlation-Linear, Causality Chi-Square tests	1
1.7	Cross validation and feature selection	1
2	**Basic Data Analytics Methods Using R**	
2.1	Introduction to R and GUI	1
2.2	Attributes and Data Types	1
2.3	Descriptive Statistics	1
2.4	Statistical methods for evaluation and Hypothesis	1
2.5	Hypothesis Testing- Difference of Means	1
2.6	Type –I, Type –II Errors, Problems	1
2.7	ANNOVA	1
2.8	Visualization – Single variable.	1
2.9	Examining multiple variables	1

3	**Models**	
3.1	Finite state machines	1
3.2	Recurrent models	1
3.3	Autoregressive models	1
3.4	Autoregressive models	1
3.5	Moving Average Models- ARIMA	1
3.6	Moving Average Models- ARMA	1
4	**Clustering**	
4.1	Cluster partitions	1
4.2	Sequential clustering, Prototype based clustering	1
4.3	Relational clustering,	1
4.4	Cluster tendency assessment, Cluster validity	1
4.5	Self-organising map (SOP).	1
4.6	Regression: Linear regression	1
4.7	Logistic Regression, Use cases	1
5	**Classification**	
5.1	Naive Bayes classifier	1
5.2	LDA	1
5.3	SVM	1

5.4	Learning Vector Quantization	1
5.5	Scalability through parallelization	1
5.6	Data parallelization, Process parallelization	1
5.7	Scaling using feature engineering	1
5.8	Cascading, Feature reduction through spatial transforms	1
5.9	ReLU nonlinearity	1
5.10	Data Augmentation, MLP, Convolutional Layer	1
5.11	Global Average Pooling, Dimensionality Reduction	1

Reference Books

1. Thomas A. Runkler, "Data Analytics - Models and Algorithms for Intelligent Data Analysis", Springer 2012.
2. Stefanos Vrochidis, Benoit Huet, Edward Chang, IoannisKompatsiaris, "Big Data Analysis for Large-Scale Multimedia Search", Wiley 2019.
3. J. O. Moreira, Andre Carvalho, Tomas Horvath, "A General Introduction to Data Analytics", Wiley 2019.
4. "Data Science and Big data Analytics, Discovering, Analyzing, Visualizing and Presenting Data" EMC Education Service.
5. Gouzhu Dong and Huan Liu, " Feature Engineering For Machine Learning and Data Analytics " CRC Press..

221ECS013	R FOR DATA SCIENCE	CATEGORY	L	T	P	CREDIT
		Program Elective 1	3	0	0	3

Preamble:

The course introduces the R programming environment and its use in Data Science. It covers data loading and organization, data exploration and cleaning, building machine learning models, statistical models, and documentation and data visualization using R. The course enables learners to use R programming in data analytics related tasks for making predictions.

Course Outcomes:

After the completion of the course the student will be able to

CO 1	Organize, explore, clean and analyse data to find relative patterns in data. (**Cognitive Knowledge Level: Apply**)
CO 2	Design different machine learning models to make predictions from data. (**Cognitive Knowledge Level: Apply**)
CO 3	Demonstrate the patterns in data using various data visualization packages. (**Cognitive Knowledge Level: Apply**)
CO4	Build prediction models for different types of data, evaluate and validate them. (**Cognitive Knowledge Level: Apply**)
CO5	Apply R programming skills to solve real-life data analytics problems. (**Cognitive Knowledge Level: Apply**)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams.

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards.

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects.

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✔		✔	✔	✔	✔	
CO 2	✔		✔	✔	✔	✔	
CO 3	✔		✔	✔	✔	✔	
CO 4	✔		✔	✔	✔	✔	
CO5	✔	✔	✔	✔	✔	✔	✔

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	60
Analyse	40
Evaluate	-
Create	-

Assignments or course projects can be used for higher level assessment of course outcomes.

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Given a vector of values, demonstrate how would you convert it into a time series object?

2. If x = c(1, 2, 3, 3, 5, 3, 2, 4, NA), what are the levels of factor(x)?

3. Write a custom function which will replace all the missing values in a vector with the mean of values.

Course Outcome 2 (CO2)

1. Demonstrate the use of any five generic functions for extracting model information.

2. Elaborate on how Anova model can be used for data analysis.

3. Describe how to fit a non-linear regression model in R.

Course Outcome 3 (CO3):

1. With the help of an example, show how to create scatter plot using R libraries.

2. Analyse the type of chart to be used when trying to demonstrate the relationship between variables/parameters.

3. Suggest scenarios where you would use a bar chart and a histogram. Justify.

Course Outcome 4 (CO4):

1. Suppose, given a dataset (x1, y1), (x2, y2), (x3, y3)…..(xn, yn) of n observation from an experiment fit a straight line y=a+bx to the given data.

2. Compare the situation where you want to compare your data distribution with another. How would you accomplish using R, the scenario where you need to check if a sample follows a normal distribution or not or if two samples are drawn from the same distribution?

3. Given a dataset, elaborate on how you would choose a suitable model for prediction. Justify your answer.

Course Outcome 5 (CO5):

1. Write an R program to read a given a dataset, clean the data, organize the data and build a suitable machine learning model to make predictions. Also, evaluate and validate the model.

2. Suppose you are asked to build a model for spam detection, discuss about the method you would follow if you would prefer using: a) supervised learning method b) an unsupervised learning approach. Analyse which one would be better.

3. Choose any real-life scenario and apply R programming skills to develop a good machine learning model for making predictions. Also, evaluate the performance of your model.

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES: 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS013

Course Name: R for Data Science

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Demonstrate the working of which function in vectors with the help of suitable examples.
2. Compare and contrast list and data frames.
3. Elaborate the concept of generalized linear model and the use of glm() function.
4. With the help of a suitable example, illustrate how accuracy of a model is evaluated.
5. Show how the different plot functions in R could be used for data visualization. **(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6. Illustrate any five vector methods with appropriate examples. **(7)**
7. Write an R code to generate an upper triangular matrix in R. Convert the same into a lower triangular matrix. **(7)**
8. Compare linear regression and logistic regression. Explain how to create a linear regression model in R. **(7)**
9. The number of awards earned by students at one high school. Predictors of the number of awards earned include the type of program in which the student was enrolled (e.g., vocational, general or academic) and the score on their final exam in math. Select a suitable regression model to perform the analysis. Also write the code for the same. **(7)**
10. Describe the common probability distribution functions in R and their applications. **(7)**
11. Justify the use of scatter plot in data analytics by quoting suitable application. **(7)**

12. Given is the 'diamonds' dataset in R which is part of the ggplot2 library. It contains prices of approximately 50000 round cut diamonds. How would you use an approach to plot a histogram that will display a type of diamonds based on the quality of cut (Ideal, Premium, Very Good, Good and Fair). **(7)**

Syllabus

	Syllabus	
Module	**Content**	**Hours**
1	**Introduction** - Reading and getting data into R, Vectors and assignment, Logical and Index vectors, Generating regular sequences, Missing values, Ordered and Unordered Factors, The function tapply() and ragged arrays, Ordered factors.	**9**
2	**Exploring and cleaning data for analysis** - Reading data from files, Data organization, Arrays and Matrices, Basics of Arrays in R, Matrix operations, Advanced Matrix operations, Additional Matrix facilities, Lists and Data frames.	**11**
3	**Building machine learning models** - Building linear models, Generalized linear models, Nonlinear least squares and maximum likelihood models.	**8**
4	**Evaluating and Validating models** - Evaluating and Validating models, Probability distributions in R, Statistical models in R.	**5**
5	**Data Visualization** - Documentation, Graphical analysis, plot() function, Displaying multivariate data, Using graphics parameters, Matrix plots, Exporting graphs, ggplot package.	**7**

Course Plan

No	**Topic**	**No. of Lectures**
1	**Introduction to R**	
1.1	Introduction to R environment, Installation of R environment and R studio.	1
1.2	Variables and datatypes in R	1
1.3	Reading and getting data into R	1
1.4	Vectors and Assignment	1
1.5	Logical and Index vectors	1
1.6	Generating regular sequences	1
1.7	Missing values, Ordered and Unordered Factors	1
1.8	The function tapply() and ragged arrays	1
1.9	Ordered factors	1
2	**Exploring and cleaning data for analysis**	

2.1	Reading data from files	1
2.2	Data organization	1
2.3	Arrays and Matrices	1
2.4	Basics of Arrays in R	1
2.5	Matrix operations	1
2.6	Advanced Matrix operations	1
2.7	Additional Matrix facilities	1
2.8	Introduction to Lists	1
2.9	Introduction to Lists	1
2.10	Data frames	1
2.11	Data frames	1
3	**Building machine learning models**	
3.1	Building linear models	1
3.2	Building linear models	1
3.3	Generalized linear models	1
3.4	Generalized linear models	1
3.5	Nonlinear least squares	1
3.6	Nonlinear least squares	1
3.7	Maximum likelihood models	1
3.8	Maximum likelihood models	1
4	**Evaluating and Validating models**	
4.1	Evaluating and Validating models	1
4.2	Probability distributions in R	1
4.3	Probability distributions in R	1
4.4	Statistical models in R	1
4.5	Statistical models in R	1
5	**Data Visualization**	
5.1	Graphical analysis, plot()	1
5.2	Displaying multivariate data	1
5.3	Using graphics parameters	1
5.4	Matrix plots	1
5.5	Exporting graphs	1
5.6	ggplot package	1
5.7	Documentation	1

References:

1. Roger D. Peng, “R Programming for Data Science”, Lean Publishing, 2015.

2. Nina Zumel, John Mount "Practical Data Science with R. Manning Publications. 2014

3. Nathan Yau, "Visualize This: The Flowingdata Guide to Design, Visualization and Statistics", Wiley, 2011.

4. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, "Mining of Massive Datasets",Cambridge University Press, 2014.

5. Tilman M. Davies, ‘The Book of R - A First Course in R Programming and Statistics’, No Starch Press, 2016.

6. Tony Ojeda, Sean Patrick Murphy, Benjarnin Bengfort. Abhijit Dasgupta. "Practical Data Science Cookbook", Packt Publishing Limited, 2014.

7. W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R", 2013

221ECS014	Data visualization with python	CATEGORY	L	T	P	CREDIT
		Program Elective 1	3	0	0	3

Preamble: This course is intended to provide basic concepts of Data Visualization. This course helps students learn visualization libraries in python and apply them to obtain and understand the underlying information. This course helps students to implement the visualization techniques to build an interactive dashboard

Course Outcomes:

After the completion of the course, the student will be able to

CO1	Analyze the need for data Visualization in Data Analytics (**Cognitive knowledge level: Apply)**
CO2	Identify the right tool for Data Visualization based on the Data Analytics Problem (**Cognitive knowledge level: Apply)**
CO3	Utilize the right visualization technique to obtain information, knowledge, and insight from a particular Dataset. (**Cognitive knowledge level: Apply)**
CO4	Implement Visualization tasks using Python (**Cognitive knowledge level: Apply)**
CO5	Apply Plotly to create plots like Bar Charts, Line Charts, Scatter Plots, and Heat Maps, and create and deploy an interactive dashboard (**Cognitive knowledge level: create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively and write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor's program

PO4: An ability to apply stream knowledge to design or develop solutions for real-world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tools to model, analyze and solve practical engineering problems.

PO6: An ability to engage in lifelong learning for the design and development related to the stream-related problems taking into consideration sustainability, societal, ethical, and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✔	✔	✔	✔	✔	✔	
CO 2	✔	✔	✔		✔	✔	
CO 3	✔	✔		✔		✔	✔
CO 4		✔	✔		✔	✔	
CO 5	✔	✔		✔		✔	✔

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	60-80%
Analyse	20-40%
Evaluate	Assignments/Project
Create	Assignments/Project

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

The evaluation shall only be based on application, analysis, or design-based questions (for both internal and end-semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer-reviewed original publications (minimum 10 publications shall be referred) : 15 marks
ii. Course based task / Seminar/ Data collection and interpretation : 15 marks
iii. Test paper (1 number) : 10 marks

Test paper shall include a minimum of 80% of the syllabus.

Course-based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation, and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such

questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long
answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks
Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions
Course Outcome 1 (CO1):

1. a. Distinguish count histogram, relative frequency histogram, cumulative frequency histogram, and density histogram
   b. For what type of data is a Histogram plot usually used?
   c. Explain the various features that you can find from Histogram plots
2. a. Explain the significance of Box plots
   b. List the information you could gain from a box plot.
3. a. What is a scatter plot? For what type of data is a scatter plot usually used?
   b. List the features that might be visible in scatterplots
   c. Choose the type of plot you would like to use if you need to demonstrate "the relationship" between variables/parameters

Course Outcome 2 (CO2)

1. Explain the different types of Data Structures available in pandas and the advantages of pandas
2. Explain the methods of choosing the right visualization tool
3. Explain the different types of data and levels of measurement and how they influence a data analyst to choose a chart for the data he would like to visualize

Course Outcome 3(CO3):

1. Discuss the different data types, NumPy can support and discover changing the type of the variable affects the data it stores.
2. Explain the grouping and aggregation methods in pandas
3. Look at the following data

```
country  continent year lifeExp pop      gdpPercap
0 Afghanistan   Asia    1952 28.801 8425333  779.445314
```

```
1 Afghanistan   Asia    1957 30.332 9240934  820.853030
2 Afghanistan   Asia    1962 31.997 10267083 853.100710
3 Afghanistan   Asia    1967 34.020 11537966 836.197138
4 Afghanistan   Asia   1972  36.088 13079460 739.981106
5 Afghanistan   Asia   1977  38.438 14880372 786.113360
6 Afghanistan   Asia   1982  39.854 12881816 978.011439
7 Afghanistan   Asia   1987  40.822 13867957 852.395945
8 Afghanistan   Asia   1992  41.674 16317921 649.341395
9 Afghanistan   Asia   1997  41.763 22227415 635.341351
```

For each year in our data, estimate the average life expectancy. Also, appraise about population and GDP.

Course Outcome 4 (CO4):

1. a. Explain the seaborn library. Does seaborn require matplotlib
   b. Explain the load dataset () method with an example
   c. Discuss creating line plot, violin plot, and histogram in seaborn
2. Discuss several attributes of the Pandas data frame object that you will frequently need while cleaning, pre-processing, or analyzing a data set.
3. a. Explain the seaborn library. Does seaborn require matplotlib
   b. Explain the load dataset () method with an example
   c. Discuss creating line plot, violin plot, and histogram in seaborn

Course Outcome 5 (CO5):

1. Use plotly to create interactive plots,
2. Create main types of plots with plotly and python
3. work on the actual DASH library from plotly to begin serving components and several plots as a web app in our browser. So that is going to be a little different than just a singular plotly plot. Instead, combine multiple things and have a full-service dashboard that is basically a Web service or Web app in your browser.
4. Create data and explore advanced and complex features of the dash- multiple inputs and outputs, interactive components, Controlling callbacks with state, and linking together grouped Plotly plots
5. Using Dash dashboard and python, create a basic Web app that will automatically look up and serve stock ticker data for you in two sets of timestamps that you get to choose.

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES: 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M. TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: Data visualization using python

Course Name: 221ECS014

Max. Marks: 60 **Duration: 2.5 Hours**

PART A (5 x 5)

Answer All Questions. Each Question Carries 5 Marks

1.	Explain the importance of Data Visualization	**5**
2.	Write a Pandas program to get the powers of array values element-wise. Note: First array elements raised to power the from the second array Sample data: {'X':[78,85,96,80,86], 'Y':[84,94,89,83,86],'Z':[86,97,96,72,83] Write the sample output	**5**
3.	Explain the methods to convert String to date.	**5**
4.	a. Create a scatterplot of 1000 random data points. b. Explain the steps to make a CSV file update automatically	**5**
5.	a. Compare plotly and matplotlib b. Define univariate distributions and graphs for visualizing univariate distributions	**5**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	(a)	Explain the different types of Data Visualization charts.	**(7)**
7.	(a)	a. Explain the data frame in Pandas	**(3)**
	(b)	Explain the following data frame methods and computations with examples i. Min and Max ii. Sum and Count iii. Mean, Median and Mode iv. Describe with Numeric values v. Describe with Object (with) Text values	**(4)**
8.	(a)	Write a Pandas program to get the first 3 rows of a given DataFrame. Sample Data Frame: exam_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'], 'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19], 'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1], 'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']} labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j'] Write the sample output	**(3)**

	(b)	What is the name of pandas library tools used to create a scatter plot matrix?	(1)
	(c)	Write a Pandas program to merge two given data frames with different columns and give a sample output Test Data: data1: key1 key2 P Q 0 K0 K0 P0 Q0 1 K0 K1 P1 Q1 2 K1 K0 P2 Q2 3 K2 K1 P3 Q3 data2: key1 key2 R S 0 K0 K0 R0 S0 1 K1 K0 R1 S1 2 K1 K0 R2 S2 3 K2 K0 R3 S3	(3)
9.	(a)	Elucidate the method of creating a Scatter Plot with several colors in Matplotlib with the help of an example	(2.5)
	(b)	Explain the method of adding a legend to a scatter plot in Matplotlib	(2)
	(c)	Explain the method of increasing the size of scatter points in Matplotlib	(2.5)
10.	(a)	Explain Time Series in pandas	(2)
	(b)	How will you create a series from dict in Python?	(2)
	(c)	Explain the following operations on Series in pandas. Build a dashboard using data from a CSV file i. Concatenating series ii. Concatenating series by indexing iii. Concatenating series by column iv. Data frame Merge () method	(3)
11.	(a)	Build a dashboard using data from a CSV file	(7)
12.	(a)	What is an interactive dashboard?	(1)
	(b)	Explain Dash Components: HTML components, core components, and Markdown, Using Help() with Dash	(3)
	(c)	Explain Single Callback for Interactivity	(3)

Syllabus

Mod	Content	Hrs
1	**PYTHON VISUALIZATION CHARTS & INTRODUCTION TO NUMPY AND PANDAS:** Data visualization, importance, advantages, Categories and tools, design principles, listing libraries of python, JavaScript, and R, Dimensions and measures, types of data, visualizing charts using python, general theory, creation, interpretation, conditions, and disadvantages- Bar chart, Pie chart, stacked area chart, Line chart, Histogram, scatter plot, regression plot, Combining Bar and Line chart, **Numpy** –Array, NaN and INF, Statistical Operations, Shape, Reshape, Ravel, Flatten, Sequence, Repetitions, and Random Numbers, Where, File Read and Write, Concatenate and Sorting, Dates, **Pandas-**Data Frame and Series, File Reading and Writing, Info, Shape, Duplicated, and Drop, Columns, NaN and Null Values, Imputation, Lambda Function,	**8**
2	**PANDAS-** Statistical functions, **Data Visualisation With Pandas-** Line Plot, Bar Plot, Stacked Plot, Histogram, Box Plot, Area and Scatter Plot, Hex and Pie Plot, Scatter Matrix and Subplots, **Series And Columns-**Selecting A Single and multiple Column, Series Methods, The powerful value_counts() method, Using plot() to visualize **Indexing And Sorting-** Set_Index Basics, set_index: The World Happiness Index Dataset, setting index with read_csv, sort_values intro, sorting by multiple columns, sorting text columns, sort_index, Sorting and Plotting!, loc, iloc, loc & iloc with Series, **Filtering Data Frames-** Filtering data frames with a Boolean series, Filtering With Comparison Operators, The Between Method, The isin() Method, Combining Conditions Using AND (&). Combining Conditions Using OR (|), Bitwise Negation, isna() and notna() Methods, Creating and Adding, dropping, And Removing Columns and rows	**8**
3	**PANDAS & MATPLOTLIB: UPDATING VALUES-**Renaming Columns and Index Labels, The replace () method, Updating Multiple Values Using loc[], Updates With loc[] and Boolean Masks, **Working With Types:-** Casting Types With astype(), Introducing the Category Type, Casting With pd.to_numeric(), dropna() and isna(), fillna(), Working With Dates And Times **Matplotlib** –Line Plot, Label, Scatter, Bar, and Hist Plots, Box Plot, Subplot, Pie Plot Text Color, Nesting,&Labeling, Bar Chart, Line plot &Scatter plot on Polar Axis, Animation Plot, **Pandas Plotting** –Changing Plot Styles, Adding Labels and Titles, rename (),Multiple plots on The Same Axes, Automatic Subplots, Manual Subplots With Pandas, Exporting Figures With savefig(), Grouping And Aggregating	**8**
4	**PANDAS & SEABORN: PANDAS:** Hierarchical Indexing in pandas, Working With Text in pandas, Pandas- Apply, Map And Applymap-, Combining Series And Dataframes- **Seaborn:-** The Helpful load_dataset() method, Seaborn Scatterplots, Line plots. The relplot() Method, Resizing Seaborn Plots: Aspect & Height, Histograms, KDE Plots, Bivariate Distribution Plots, Rugplots, The Amazing displot() Method. **Seaborn Categorical Plots-** Countplot, Strip & Swarm Plots, Boxplots, Boxenplots, Violinplots, Barplots, The Big Boy Catplot Method, **Controlling Seaborn Aesthetics-** Changing Seaborn Themes, Customizing Styles with set_style(), Altering Spines With despine(), Changing Color Palettes	**6**
5	**PYTHON DASHBOARDS WITH PLOTLY AND DASH:** PLOTLY BASICS-	**10**

	Plots, Line Charts, Bar Charts, Bubble Plots, Box Plots, Histograms, Distplots,, Heatmaps, DASH BASICS –dash layouts, - styling, Converting Simple Plotly Plot to Dashboard with Dash, create a simple dashboard, **Dash Components,** HTML Components, Core Components, Markdown with Dash, Using Help() with Dash INTERACTIVE COMPONENTS- Single Callbacks for Interactivity, Dash Callbacks for Graphs, Multiple Inputs &Outputs, Controlling Callbacks with Dash State, INTERACTING WITH VISUALISATION -Hover Over Data, Click Data, Selection Data, Updating Graphs on Interactions- Project: Building an Interactive dashboard with Plotly and Dash	

Course Plan

S.NO	TOPIC	NO. OF LECTURES
MODULE 1 - PYTHON VISUALIZATION CHARTS & INTRODUCTION TO NUMPY AND PANDAS: 8 hours		
1.1	Data visualization, importance, advantages, Categories and tools, design principles, listing libraries of python, JavaScript, and R, Dimensions and measures, types of data,	1
1.2	visualizing charts using python, general theory, creation, interpretation, conditions, and disadvantages- Bar chart, Pie chart,	1
1.3	Stacked area chart, Line chart	1
1.4	Histogram, scatter plot	1
1.5	Regression plot, Combining Bar and Line chart,	1
1.6	**NUMPY** –Array, NaN and INF, Statistical Operations, Shape, Reshape, Ravel, Flatten, Sequence, Repetitions, and Random Numbers, Where, File Read and Write, Concatenate and Sorting, Dates,	1
1.7	**PANDAS-**Data Frame and Series, File Reading and Writing, Info, Shape, Duplicated, and Drop, Columns,	1
1.8	NaN and Null Values, Imputation, Lambda Function,	1
MODULE 2- PANDAS- 8 hours		
2.1	**PANDAS-** Statistical functions	1
2.2	**Data Visualisation With Pandas**- Line Plot, Bar Plot, Stacked Plot, Histogram, Box Plot, Area and Scatter Plot, Hex and Pie Plot, Scatter Matrix and Subplots,	1
2.3	**Series And Columns-**Selecting A Single and multiple Column, Series Methods, The powerful value_counts() method, Using plot() to visualize	1
2.4	Selecting Multiple Columns, The powerful value_counts() method, Using plot() to visualize!, EXERCISE: Series & Plotting	1
2.5	**Indexing And Sorting-** Set_Index Basics, set_index: The World Happiness Index Dataset, setting index with read_csv, sort_values intro, sorting by multiple columns, sorting text columns, sort_index, Sorting and Plotting, loc, iloc, loc & iloc with Series	1

2.6	**Filtering Data Frames-** Filtering data frames with a Boolean series, Filtering With Comparison Operators, The Between Method, The isin() Method,	1
2.7	Combining Conditions Using AND (&). Combining Conditions Using OR (|), Bitwise Negation, isna() and notna() Methods,	1
2.8	Creating and Adding, dropping, And Removing Columns and rows	1
	MODULE 3- PANDAS & MATPLOTLIB- 8 hours	
3.1	**Updating Values**-Renaming Columns and Index Labels, The replace () method, Updating Multiple Values Using loc[], Updates With loc[] and Boolean Masks	1
3.2	**Working With Types:-** Casting Types With astype(), Introducing the Category Type, Casting With pd.to_numeric(), dropna() and isna(), fillna(),	1
3.3	Working With Dates And Times	1
3.4	**Matplotlib** –Line Plot, Label, Scatter, Bar, and Hist Plots, Box Plot, Subplot, Pie Plot Text Color, Nesting,&Labeling,	1
3.5	Bar Chart, Line plot &Scatter plot on Polar Axis, Animation Plot,	1
3.6	**Pandas Plotting**–Changing Plot Styles, Adding Labels and Titles, rename(), Multiple plots on The Same Axes, Automatic Subplots,	1
3.7	Manual Subplots With Pandas, Exporting Figures With savefig(),	1
3.8	Grouping And Aggregating	1
	MODULE 4- PANDAS & SEABORN- 6 hours.	
4.1	**PANDAS:** Hierarchical Indexing in pandas	1
4.2	Working With Text in pandas	1
4.3	Pandas- Apply, Map And Applymap	1
4.4	Combining Series And Dataframes-	1
4.5	**SEABORN:-** The Helpful load_dataset() method, Seaborn Scatterplots, Line plots. The relplot() Method, Resizing Seaborn Plots: Aspect & Height, Histograms, KDE Plots, Bivariate Distribution Plots, Rugplots, The Amazing displot() Method	1
4.6	**SEABORN CATEGORICAL PLOTS-** Countplot, Strip & Swarm Plots, Boxplots, Boxenplots, Violinplots, Barplots, The Big Boy Catplot Method, **CONTROLLING SEABORN AESTHETICS-** Changing Seaborn Themes, Customizing Styles with set_style(), Altering Spines With despine(), Changing Color Palettes	1
	MODULE 5-PYTHON DASHBOARDS WITH PLOTLY AND DASH=10 hours	
5.1	**PLOTLY BASICS-** Plots, Line Charts	1
5.2	Bar Charts, Bubble Plots, Box Plots	1
5.3	Histograms, Distplots,, Heatmaps,	1
5.4	**Dash Basics** –dash layouts, - styling, Converting Simple Plotly Plot to Dashboard with Dash, creating a simple dashboard, **Dash Components,** HTML	1

	Components, Core Components, Markdown with Dash, Using Help() with Dash	
5.5	**Interactive Components-** Single Callbacks for Interactivity, Dash Callbacks for Graphs	1
5.6	Multiple Inputs &Outputs, Controlling Callbacks with Dash State,	1
5.7	INTERACTING WITH VISUALISATION -Hover Over Data, Click Data	1
5.8	Selection Data, Updating Graphs on Interactions	1
5.9	Project: Building an Interactive dashboard with Plotly and Dash	1
5.10	Project: Building an Interactive dashboard with Plotly and Dash	1

Reference Books

TEXTBOOKS:

1. Dr Ossama Embarak, ‘Data Analysis and Visualisation Using python- Analyze Data to Create Visualizations for BI Systems -Apress
2. Kirthi Raman, “Mastering Python Data Visualization”, Packet Publishing Ltd, UK
3. Elias Dabbas, Interactive Dashboards and Data Apps with Plotly and Dash: Harness the power of a fully-fledged frontend web framework in Python ˚no JavaScript required, ISBN: 9781800568914
4. Fabio Nelli, “Python Data Analytics with pandas, NumPy, Matplotlib”, Second Edition, Apress
5. Matt Harrison and Michael Prentiss, “Learning the Pandas Library- Python tools for Data Munging, Data Analysis, and Visualisation”
6. Daniel Y. Chen, “Pandas for Everyone: Python Data Analysis, 1e Paperback”

221ECS015	**REINFORCEMENT LEARNING**	**CATEGORY**	**L**	**T**	**P**	**CREDIT**
		Program Elective 1	3	0	0	3

Preamble:

This course provides the basic concepts and advanced techniques in reinforcement learning. This course covers policies and value functions, Q-learning, function approximation and policy optimization methods. This course helps the learners to acquire key skills required to develop applications in the exciting area of reinforcement learning.

Course Outcomes:

After the completion of the course the student will be able to

CO 1	Utilize the key features of reinforcement learning developing machine learning applications. (**Cognitive knowledge level: Apply**)
CO 2	Explore suitable learning tasks to which RL techniques can be applied. (**Cognitive knowledge level: Apply**)
CO 3	Apply core principles behind the RL, including policies, value functions, deriving Bellman equations. (**Cognitive knowledge level: Apply**)
CO 4	Implement and analyze approximate solutions (**Cognitive knowledge level: Analyze**)
CO 5	Analyze current advanced techniques and applications in RL (**Cognitive knowledge level: Analyze**)
CO 6	Practice using popular open-source library for implementing RL algorithms. (**Cognitive knowledge level: Apply**)

Program Outcomes (POs)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1			✓		✓	✓	
CO 2			✓		✓	✓	
CO 3			✓		✓	✓	
CO 4			✓	✓	✓	✓	
CO 5			✓		✓	✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	60-80

Analyse	20-40
Evaluate	Can be evaluated using Mini projects/assignments
Create	Can be evaluated using Mini projects/assignments

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation: 15 marks

iii. Test paper (1 number): 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. List the elements of reinforcement learning.

2. Explain multi-armed bandit problems.

3. Explain action-value methods.

Course Outcome 2 (CO2)

1. Give examples of Markovian and non-Markovian environments.

2. What are the advantages and disadvantages of value methods vs policy methods?

3. Imagine that the rewards are at most 1 everywhere. What is the maximum value that the discounted return can attain ? Why?

Course Outcome 3(CO3):

1. Describe Monte Carlo prediction, estimation and control.

2. 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?

3. With respect to the expected Sarsa algorithm, is exploration required as it is in the normal Sarsa and Q-learning algorithms? Justify.

Course Outcome 4 (CO4):

1. Describe value function approximation.
2. Justify the use of deep neural networks in nonlinear function approximation.
3. In the context of λ-return, differentiate between simple update and compound update.

Course Outcome 5 (CO5):

1. State the policy gradient theorem.
2. Compare two versions of actor-critic algorithms.
3. What are the basic elements of OpenAI Gym environment.

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES: 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M. TECH DEGREE EXAMINATION, MONTH & YEAR

Model Question Paper

Course Code: 221ECS015

Course Name: Reinforcement Learning

Max. Marks: 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Describe briefly the elements of reinforcement learning.
2. Write down the Bellman expectation equation for state-value functions.
3. Draw the backup diagram for 2-step Q-learning. Write the corresponding learning rule for 2-step Q-learning.
4. Compare and contrast any two linear methods used for function approximation.
5. State the policy gradient theorem and its applications. **(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	Describe the principles behind incremental implementation of computations for estimated action values. Give a simple algorithm for the same.	**(7)**
7.	Distinguish between policy iteration and value iteration. Give relevant algorithms.	**(7)**
8.	Why is Q-learning considered an off-policy control method?	**(7)**
9.	Demonstrate how polynomial function approximation can be help in a reinforcement learning problem having 3 dimensional states.	**(7)**
10.	Derive the REINFORCE policy-gradient learning algorithm.	**(7)**
11.	Prove that the approximation for the λ-return becomes exact if the approximate value function does not change.	**(7)**
12.	Provide pseudocode of the actor-critic algorithm that uses eligibility traces.	**(7)**

Syllabus

Module 1: Introduction to reinforcement learning (8 hours)
Introduction to RL, Examples, Elements of RL, Multi-armed bandit problems, Action-value methods, The ten-armed testbed, Incremental implementation
Module 2: Policies and value functions (8 hours)
Markov Decision Process, Goals and rewards, Returns and episodes, Policies and value functions, Policy evaluation, Policy improvement, Policy iteration, Value iteration
Module 3: Q learning (8 hours)
Monte Carlo prediction, estimation and control, TD prediction, Sarsa, Q-learning, n-step TD prediction, n-step Sarsa, n-step Off-policy learning, Dyna

Module 4: Function approximation (8 hours)
Value function approximation, Stochastic gradient methods, Linear methods, Non-linear function approximation, Episodic semi-gradient control, Semi-gradient n-step Sarsa, The λ return, TD(λ)
Module 5: Policy approximation (8 hours)
Policy approximation, Policy gradient theorem, REINFORCE algorithm, Actor-Critic methods, Trust-Region Policy Optimization, Proximal Policy Optimization, Introduction to OpenAI Gym

Course Plan

No	Topic	**Number of Hours (39 Hours)**
1	**Introduction to reinforcement learning (7 hours)**	
1.1	Introduction to RL.	1
1.2	Examples	1
1.3	Elements of RL	1
1.4	Multi-armed bandit problems	1
1.5	Action-value methods	1
1.6	The ten-armed testbed	1
1.7	Incremental implementation	1
2	**Policies and value functions (8 hours)**	
2.1	Markov Decision Process	1

2.2	Goals and rewards	1
2.3	Returns and episodes	1
2.4	Policies and value functions	1
2.5	Policy evaluation	1
2.6	Policy improvement	1
2.7	Policy iteration	1
2.8	Value iteration	1
3	**Q learning (8 hours)**	
3.1	Monte Carlo prediction, estimation and control	1
3.2	TD prediction	1
3.3	Sarsa	1
3.4	Q-learning	1
3.5	n-step TD prediction	1
3.6	n-step Sarsa	1
3.7	n-step Off-policy learning	1
3.8	Dyna	1
4	**Function approximation (8 hours)**	
4.1	Value function approximation	1
4.2	Stochastic gradient methods	1

4.3	Linear methods	1
4.4	Non-linear function approximation	1
4.5	Episodic semi-gradient control	1
4.6	Semi-gradient n-step Sarsa	1
4.7	The λ-return	1
4.8	TD(λ)	1
5	**Policy methods (8 hours)**	
5.1	Policy approximation	1
5.2	Policy gradient theorem	1
5.3	REINFORCE algorithm	1
5.4	Actor-Critic methods	1
5.5	Trust-Region Policy Optimization	1
5.6	Proximal Policy Optimization	1
5.7	Introduction to Open AI Gym	1
5.8	Introduction to Open AI Gym	1

Reference Books

1. Reinforcement learning: an introduction, Richard S. Sutton and Andrew G. Barto, Second edition, MIT Press, 2018.

2. Algorithms for Reinforcement Learning. C. Szepesvari. Morgan and Claypool Publishers, 2010

3. Reinforcement Learning: State-of-the-Art. M. Wiering and M. van Otterlo. Springer, 2012

Reference Papers

1. John Schulman, Sergey Levine, Philipp Moritz, Michael Jordan, and Pieter Abbeel. 2015. Trust region policy optimization. In Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (ICML'15). JMLR.org, 1889– 1897.

2. Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg limov. "Proximal policy optimization algorithms." *arXiv preprint arXiv:1707.06347* (2017)

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
221ECS016	Computational Linguistics	Program Elective 1	3	0	0	3

Preamble:

This course introduces the fundamentals of Language processing from a computational viewpoint. This course covers Language models, Computational Phonology and Morphology Unification, Semantics and knowledge representation and Pragmatics. It helps the student to apply NLP tasks such as POST, WSD, and modeling of languages.

Program Outcomes

Graduates of this program will be able to demonstrate the following attributes.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams.

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards.

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tools to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects.

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Course Outcomes: The COs shown are only indicative. For each course, there can be 4 to 6 COs.

After the completion of the course the student will be able to

CO 1	Apply Probabilistic Models of Pronunciation and Spelling (Cognitive Knowledge Level: Apply)
CO 2	Apply the different methods for Parsing with Context-Free Grammars for English (Cognitive Knowledge Level: Apply)
CO 3	Apply basic concepts for Probabilistic Context-Free Grammars (Cognitive Knowledge Level: Apply)
CO 4	Describe Unification of Feature Structures (Cognitive Knowledge Level: Understand)
CO 5	Apply the key concepts Word Sense Disambiguation and Information Retrieval (Cognitive Knowledge Level: Apply)
CO 6	Develop an application that uses Natural Language Generation concepts (Cognitive Knowledge Level: Apply)

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓		✓	✓	
CO 2	✓		✓		✓	✓	
CO 3	✓		✓		✓	✓	
CO 4			✓		✓	✓	
CO 5	✓		✓		✓	✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	50-60
Analyse	30-40
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Test 1	Test 2	Assignments	Total
15	15	10	40

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 5 questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B contains 7 full questions from each module of which students should answer any 5 full questions. Each question can have a maximum 2 subdivisions and carries 7 marks.

Model Question paper

APJ ABDULKALAM TECHNOLOGICAL UNIVERSITY		
Q. P. Code :		**Name:** .... **Reg. No:** ....
FIRST SEMESTER M.TECH. DEGREE EXAMINATION		
Branch: Computer Science and Engineering		
221ECS016 Computational Linguistics		
Time: 2.5 hours		**Max. Marks: 60**
Answer all 5 questions.		

Q. No.	**Part A** **Answer all 5 questions.**	**Marks**
1.	Compare inflectional and derivational morphology.	**5**
2.	What are the different types of single-error misspellings give examples for each	**5**

3.	S→ NP VP NP→ Det N VP→ V NP N→ flight | meal V→ includes Det→ the | a Parse the sentence ***“the flight includes a meal”*** Using CYK algorithm	**5**
4.	Discuss the different feature structures associated with grammar. Explain with examples	**5**
5.	Explain polysemy with example.	**5**
Q. No.	**Part B** **Answer any 5 questions**	**Marks**
6.	Write Chomsky and Halle notation for the following rules i.Keep the first letter of the name,and drop all occurrences of a,e,i,o,u,w,y ii.Replace any sequence of identical numbers with a single number (ie., 333→ 3)	**7**
7.	Write a program(python is sufficient) to compute unsmoothed bigrams count	**7**
8.	Write a short note on Human parsing	**7**

9.	Draw the DAGs corresponding to the AVMs given in Examples 	**7**
10.	Justify the need of Word Sense Disambiguation? Explain supervised method of WSD in detail	**7**
11.	Explain the selectional restriction-based disambiguation and its limitations	**7**
12.	Describe the different feature structures associated with grammar. Explain with examples	**7**

Syllabus and Corse Plan (For 3 credit courses, the content can be for 40 hrs and for 2 credit courses, the content can be for 26 hrs. The audit course in third semester can have content for 30 hours).

No	Topic	No. of Lectures
1	**Introduction (8 Hours)**	
1.1	Words-Regular Expressions	1
1.2	Automata	1
1.3	Morphology	1
1.4	Finite-State Transducers	1

1.5	Computational Phonology	1
1.6	Pronunciation Modeling	1
1.7	Probabilistic Models of Pronunciation	1
1.8	Probabilistic Models of Spelling	1
2	**Syntax (6 Hours)**	
2.1	N-gram models	1
2.2	N-gram models of Syntax-Word Classes	1
2.3	Part- of-Speech Tagging	1
2.4	Context-Free Grammars for English	1
2.5	Parsing	1
2.6	Parsing with Context-Free Grammars	1
3	**Probabilistic Context-Free Grammars (7 Hours)**	
3.1	Probabilistic CYK	1
3.2	Parsing of PCFGs	1
3.3	Learning PCFG Probabilities	1
3.4	Problems with PCFGs	1
3.5	Probabilistic Lexicalized CFGs	1
3.6	Dependency Grammars	1
3.7	Human Parsing	1
4	**Unification of Feature Structures (6 Hours)**	

4.1	Feature Structures in the Grammar	1
4.2	Agreement-Head Features	1
4.3	Subcategorization	1
4.4	Long Distance Dependencies	1
4.5	Implementing Unification	1
4.6	Unification Data Structures-	1
5	**Semantics and Pragmatics (7 Hours)**	
5.1	Representing Meaning	1
5.2	Semantic Analysis	1
5.3	Lexical Semantics	1
5.4	Word Sense Disambiguation	1
5.5	Information Retrieval	1
5.6	Natural Language Generation	1
5.7	Machine Translation	1

Reference Books

1. Jurafsky, D. and J. H. Martin, Speech and language processing:
2. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition,Prentice-Hall, 2000.
3. Charniak, E.: Statistical Language Learning. The MIT Press.
4. J. Allen: Natural Language Understanding. Benjamin/Cummins.

221ECS004	COMPUTATIONAL INTELLIGENCE	CATEGORY	L	T	P	CREDIT
		Program Elective 1	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 effective and efficient Computational Intelligence solutions to problems for which a direct solution is impractical or unknown. This course covers concepts of fuzzy logic, genetic algorithms, and swarm optimization techniques. The learners will be able to provide Fuzzy and AI –based solutions to real world problems.

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

CO 1	Apply fuzzy logic to handle uncertainty and solve engineering problems. **(Cognitive Knowledge Level: Apply)**
CO 2	Apply Fuzzy Logic Inference methods in building intelligent machines. **(Cognitive Knowledge Level: Apply)**
CO 3	Design genetic algorithms for optimized solutions in engineering problems. **(Cognitive Knowledge Level: Analyze)**
CO 4	Analyze the problem scenarios and apply Ant colony system to solve real optimization problems. **(Cognitive Knowledge Level: Analyze)**
CO 5	Apply PSO algorithm to solve real world problems. **(Cognitive Knowledge Level: Apply)**
CO6	Design, develop and implement solutions based on computational intelligence concepts and techniques. **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and developmentwork in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1				✔		✔	
CO 2	✔		✔	✔	✔	✔	
CO 3	✔		✔	✔	✔	✔	
CO 4	✔		✔	✔	✔	✔	
CO 5	✔		✔	✔	✔	✔	
CO 6	✔	✔	✔	✔	✔	✔	✔

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	70%-80%
Analyze	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.
Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Let V = {A, B, C, D} be the set of four kinds of vitamins, F = {f_1, f_2, f_3} be three kinds of fruits containing the vitamins to various extents, and D = {d_1, d_2, d_3} be the set of three diseases that are caused by deficiency of these vitamins. Vitamin contents of the fruits are expressed with the help of the fuzzy relation R over F x V, and the extent of which diseases are caused the deficiency of these vitamins is given by the fuzzy relation S over V x D. Relations R and S are given below

$R = [0.5\ 0.2 \quad 0.2\ 0.7 \quad 0.4\ 0.4 \quad 0.1\ 0.1\ 0.4\ 0.3\ 0.8\ 0.1\ ]S$
$= [0.3\ 0.5\ 0.1\ 0.8\ \ 0.7\ 0.4\ \ 0.9\ 0.1\ 0.5\ 0.5\ \ 0.2\ 0.3\ \ ]$

Find the correlation between the amount of certain fruit that should be taken while suffering from a disease.

Course Outcome 2 (CO2):

1. In mechanics, the energy of a moving body is called kinetic energy. Suppose we model mass and velocity as inputs to a moving body and energy as output. Observe the system for a while and the following rule is deduced.

   IF x is small and y is high

   THEN z is medium

   The graphical representation of rule is given below. Let the inputs given are 0.35kg and 55m/s. What will the output using Mamdani inference? Any defuzzification method can be used to obtain the crisp single output.

Course Outcome 3(CO3):

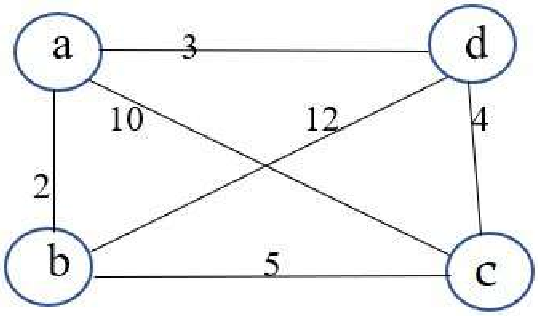
1. Describe how Roulette wheel is used for selection. Draw the Roulette wheel for six chromosomes corresponding to the table given below.

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

Course Outcome 4 (CO4):

1. Consider an Ant Colony System based on 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 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



2. Six jobs go first on machine A, then on machine B, and finally on machine C. The order of the completion of the jobs in the three machines is given in Table

Jobs	Processing time(hr)		
	Machine A	Machine B	Machine C
1	8	3	8
2	3	4	7
3	7	5	6
4	2	2	9
5	5	1	10
6	1	6	9

Find the sequence of jobs that minimizes the time required to complete the jobs using the ACS model.

Course Outcome 5 (CO5):

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) = *); x2* = (7,3); x3* = (5,6)

   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 6 (CO6):

1. Implement travelling salesman problem using appropriate optimization technique.

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES: 5**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M. TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS004

Course Name: Computational Intelligence

Max. Marks: 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Consider the set of Colours A= {Blue, Red, Orange, Yellow, Green}, Attributes B = {Bright, Warmth, Dullness}, Feelings C= {Unpleasant, happiness, Angry}. Given R and S where R is the relationship between colours and their attributes and S is the relationship between colour attributes and feelings created. Find the relationship Q between colours and feelings created **(5)**

R	Bright	Warmth	Dullness
Blue	0.8	0.6	0.4
Red	0.8	0.8	0.2
Orange	0.5	0.7	0.2
Yellow	0.3	0.6	0.5
Green	0.8	0.6	0.4

S	Unpleasant	Happiness	Angry
Bright	0.2	0.8	0.6
Warmth	0.4	0.7	0.8
Dullness	0.8	0.3	0.6

2. Develop a membership function for "Tall". Based on that devise membership function for "Very Tall". Explain how it is done **(5)**

3. Mention the importance of objective (fitness) function in genetic algorithm **(5)**

4. Describe how pheromone is updated. What is elitist / elastic ants ? Are they useful in this scenario? **(5)**

5. What is the significance of pbest and gbest particles in solving problems with particle swarm optimization? **(5)**

Part B

(Answer any five questions. Each question carries 7 marks)

6. (a) Consider the set of fruits F = {Apple, Orange, Lemon, Strawberry, Pineapple}. **(3)**

Let sweet fruits B= $\{\frac{0.8}{Apple} + \frac{0.6}{Orange} + \frac{0.2}{Lemon} + \frac{0.4}{Strawberry} + \frac{0.7}{Pineapple}\}$ and

Sour Fruits F= $\{\frac{0.6}{Apple} + \frac{08}{Orange} + \frac{0.9}{Lemon} + \frac{0.7}{Strawberry} + \frac{0.5}{Pineapple}\}$

Find Fruits that are Sweet or Sour, Sweet but not Sour, Sweet and Sour

(b) Consider two fuzzy Sets given by **(4)**

$$P = \left\{\frac{0.9}{short} + \frac{0.3}{medium} + \frac{0.5}{tall}\right\}$$

$$Q = \left\{\frac{0.7}{positive} + \frac{0.4}{zero} + \frac{0.8}{negative}\right\}$$

Find the fuzzy relation for the Cartesian product of P and Q i.e, R=P x Q .

Introduce a fuzzy set T given by

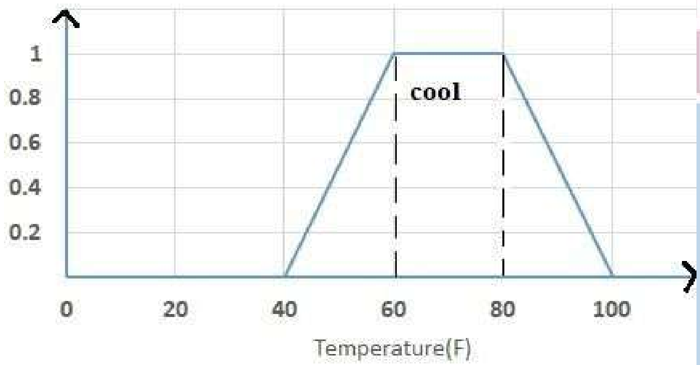
$$T = \left\{\frac{0.9}{short} + \frac{0.3}{medium} + \frac{0.6}{tall}\right\}$$

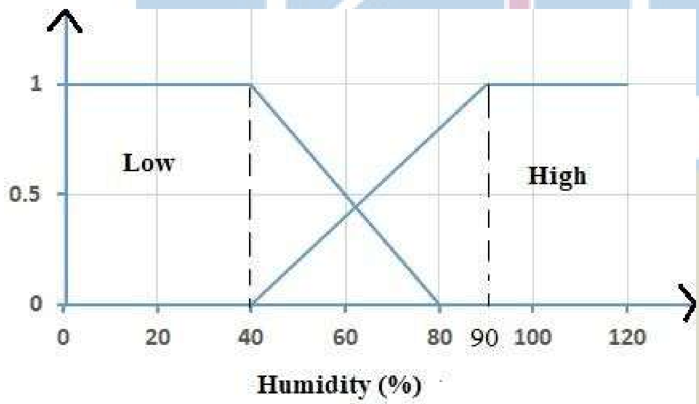
and Find T o R using max-min composition

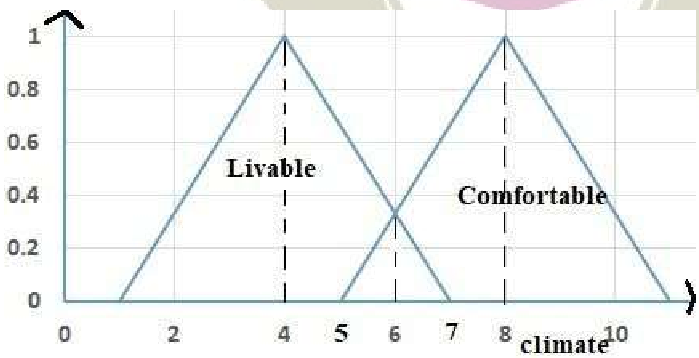
7. Consider a Fuzzy Inference System for checking climate comfortability of human beings for long time living. The system accepts two inputs – temperature and humidity. The rules and membership functions of FIS is given below. Using Mamdani inference and center of sum, calculate output when the temperature is 50Fahrenheit and humidity is 50%. (7)

Rule 1: IF temperature is cool and humidity is low, THEN climate is comfortable.

Rule 2: IF temperature is cool and humidity is high, THEN climate is livable







The fuzzy sets "Easy Question Paper" and their corresponding "Student Performance" are given below

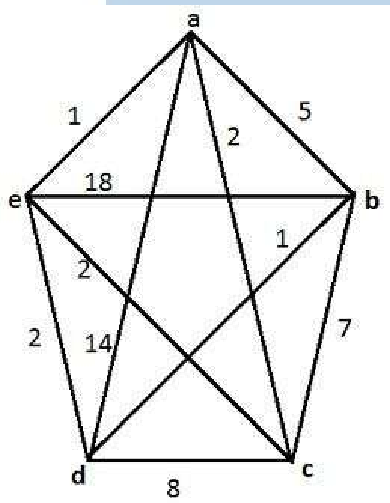
$$\text{Easy\_QP} = \left\{ \frac{0.8}{1} + \frac{0.2}{2} + \frac{0.6}{3} + \frac{0.7}{4} \right\}$$

$$\text{Stud\_Perf} = \left\{ \frac{0.3}{a} + \frac{0.4}{b} + \frac{0.8}{c} + \frac{0.9}{d} \frac{0.8}{1} + \frac{0.2}{2} + \frac{0.6}{3} + \frac{0.8}{4} + \frac{0.7}{5} \right\}$$

Find the performance of students c and d for the question paper "Somewhat Easy"

$$\text{Somewhat\_Easy} = \left\{ \frac{0.7}{1} + \frac{0.3}{2} + \frac{0.5}{3} + \frac{0.6}{4} \right\}$$

8.		Explain any procedure to map a solution to the corresponding chromosome and vice versa in genetic algorithms. Also illustrate it with an example	**(7)**
9.		Describe two methods used to select individuals from a population for the mating pool in Genetic Algorithms	**(7)**
10.	(a)	Consider the TSP with the following edge costs. Given the evaporation factor ρ =0.02 and initial pheromone at all edges T_{ij}=100	**(1)**



What is the cost of best tour?

(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<a,c> when 10 ants uses the edges <a, c>, using the following models: **(6)**

i. Ant Density Model (Constant Q=10)

ii. Ant Quantity Model(Constant Q=100)

where Q is the constant related to the pheromone updation.

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

12. 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: **(7)**

Position of particles: x1 = (4,4); x2 = (8,3); x3 = (6,7)

Individual best positions: x1* = (4,4); x2* = (7,3); x3* = (5,6)

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?

Syllabus

Module 1: Fuzzy Logic

Crisp sets vs fuzzy sets- Operations and properties of Fuzzy sets. Membership functions - Linguistic variables. Operations on fuzzy sets- Fuzzy laws- Operations on fuzzy relations, Fuzzy composition- Max- min, Max – product. Alpha-cut representation.

Module 2: Fuzzy Systems

Fuzzy Reasoning – GMP and GMT. Fuzzy Inference System: Defuzzification methods - Fuzzy Controllers -Mamdani FIS, Larsen Model

Module 3: Genetic Algorithms

Introduction to Genetic Algorithms – Theoretical foundation - GA encoding, decoding - GA operations – Elitism – GA parameters – Convergence. Multi-objective Genetic Algorithm – Pareto Ranking.

Module 4: Ant Colony Systems

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

Module 5: Particle Swarm Optimization

Basic Model - Global Best PSO- Local Best PSO- Comparison of 'gbest' to 'lbest'- PSO Algorithm Parameters- Problem Formulation of PSO algorithm- Working. Rate of convergence improvements -Velocity clamping- Inertia weight- Constriction Coefficient- Boundary Conditions- Guaranteed Convergence PSO- Initialization, Stopping Criteria, Iteration Terms and Function Evaluation.

Course Plan

No	Topic	No. of Lectures (40)
1	**Module 1: Fuzzy Logic**	**9**
1.1	Crisp sets vs fuzzy sets, Operations and properties of Fuzzy sets	1
1.2	Membership functions	1
1.3	Linguistic Variables	1
1.4	Operations on fuzzy sets	1
1.5	Fuzzy laws	1
1.6	Operations on fuzzy relations	1
1.7	Fuzzy Composition- Max- min	1
1.8	Fuzzy Composition – Max- Product	1
1.9	Alpha-cut representation	1
2	**Module 2: Fuzzy Systems**	**7**
2.1	Fuzzy Reasoning – GMP	1
2.2	Fuzzy Reasoning –GMT	1
2.3	Fuzzy Inference System	1
2.4	Defuzzification methods	1
2.5	Fuzzy Controllers	1
2.6	Mamdani Model	1
2.7	Larsen Model	1

3	**Module 3: Genetic Algorithms**	**7**
3.1	Introduction to Genetic algorithm	1
3.2	Theoretical foundation	1
3.3	GA encoding - decoding	1
3.4	GA operations	1
3.5	Elitism, GA parameters, Convergence of GA	1
3.6	Multi – objective Genetic Algorithm	1
3.7	Pareto Ranking	1
4	**Module 4: Ant Colony Systems**	**8**
4.1	Swarm intelligent systems	1
4.2	Background	1
4.3	Ant colony systems – biological systems	1
4.4	Development of the ant colony system	1
4.5	Working	1
4.6	Pheromone updating	1
4.7	Types of ant systems	1
4.8	ACO algorithms for TSP	1
5	**Module 5: Particle Swarm Optimization**	**9**
5.1	Basic Model	1
5.2	Global Best PSO	1
5.3	Local Best PSO, Comparison of 'gbest' to 'lbest'	1
5.4	PSO Algorithm Parameters	1
5.5	Problem Formulation	1
5.6	Working	1
5.7	Rate of convergence improvements – velocity clamping	1
5.8	Inertia-weight - Constriction Coefficient- Boundary Conditions	1
5.9	Initialization, Stopping Criteria, Iteration Terms and Function Evaluation	1

References

1. Samir Roy, Udit Chakraborty, Introduction to Soft Computing Neuro- Fuzzy Genetic Algorithms, Pearson, 2013

2. N.P. Padhy, Artificial Intelligence and Intelligent systems, Oxford Press, New Delhi, 2005.
3. Xin-She Yang School of Science and Technology, Middlesex University London, Nature-Inspired Optimization Algorithms, Elsevier, First edition, 2014
4. Satyobroto Talukder, Blekinge Institute of Technology, Mathematical Modelling and Applications of Particle Swarm Optimization, February 2011
5. Mitchell Melanie, An Introduction to Genetic Algorithm, Prentice Hall,1998
6. Andries Engelbrecht, Computational Intelligence: An Introduction, Wiley, 2007
7. Marco Dorigo and Thomas Stutzle, “Ant Colony optimization”, Prentice Hall of India, New Delhi 2005

CODE 221ECS017	ADVANCED DATABASE	CATEGORY	L	T	P	CREDIT
		PROGRAM ELECTIVE 2	3	0	0	3

Preamble:

This course provides an exposure to the concepts and techniques in advanced database management. The topics covered in this course includes Relational Model –Conceptual Model and Schema Design, Strategies regarding query processing and optimization, Distributed system architecture, Semi-structured data handling and modern data management techniques. This course helps the learners to develop applications that manage data efficiently with the help of suitable data models and techniques.

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

CO 1	Make use of the concepts in relational database systems including: data modelsrelational algebra, ER features, and the different normalization techniques to relational models. **(Cognitive Knowledge Level: Apply)**
CO 2	Illustrate the basic database storage, file organization, database accessing and indexing **(Cognitive Knowledge Level: Apply)**
CO 3	Identify various measures of query processing and optimization. **(Cognitive Knowledge Level: Apply)**
CO 4	Analyze implementation aspects of distributed system on database architecture. **(Cognitive Knowledge Level: Analyze)**
CO 5	Make use of semi structured data, XML and XML queries for data management. **(Cognitive Knowledge Level: Apply)**
CO 6	Design, Develop, and Implement innovative ideas on advanced database concepts and techniques. **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✔		✔	✔	✔	✔	
CO 2	✔		✔	✔		✔	
CO 3	✔		✔	✔		✔	
CO 4	✔		✔	✔		✔	
CO 5	✔		✔	✔	✔	✔	
CO 6	✔	✔	✔	✔	✔	✔	✔

Assessment Pattern

Bloom's Category	End Semester Examination

Apply	70%-80%
Analyze	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

i. **Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks**

ii. **Course based task / Seminar/ Data collection and interpretation : 15 marks**

iii. **Test paper (1 number) : 10 marks**

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

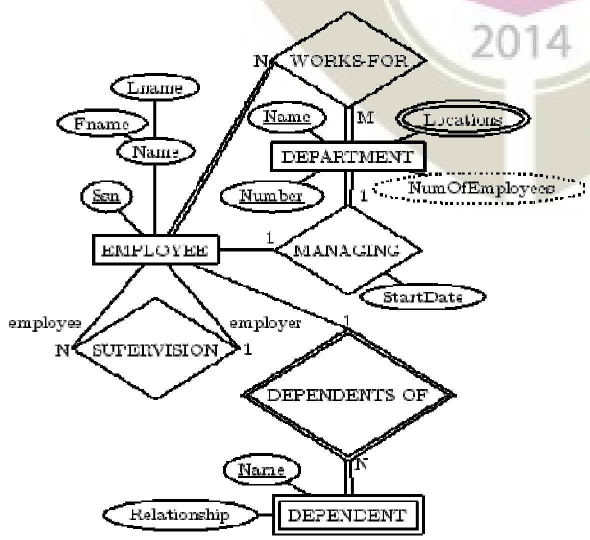
i. How does a query tree represent a relational algebra expression? Draw the initial query tree, apply heuristic rules and obtain an optimized query expression for the following SQL query.

SELECT S.SName, C.CName, G.grade

FROM Student S, Courses C, Faculty F, Grades G

WHERE S.Sid = G.Sid AND C.Cid = G.CId AND F.Fid =G.Fid AND G.grade >= 7.0 AND F.DName = 'CSE'

ii. Consider the ER diagram shown below. Identify the minimum set of relations required to map to a relational model. Identify foreign keys and primary keys. Draw a schema diagram showing all relations



iii.

Course Outcome2 (CO2):

1. Differentiate between fixed length records and variable length records.
2. A file has r=36000 STUDENT records of fixed length. Each record has the following fields: Name (25 bytes), SSN (8 bytes), Address(35 bytes), Phone(10 bytes), Date of Birth (8 bytes), Sex(1 bytes), Class code (3 bytes).
   i. Calculate the record size of R
   ii. Calculate the blocking factor bfr and number of file blocks b, assuming an unspanned organization
   iii. Calculate the average time it takes to find a record doing a linear search on the file (assume blocks are stored continuously)
   iv. Assume the file ordered by SSN, by doing binary search, calculate time it takes.
   v. Assume a primary index is created with key as SSN and the data pointer needs 8 bytes, find the number blocks required to keep the primary index. Also find the average time required to find a record using the index.

Course Outcome3 (CO3):

i. Let *s* be selection cardinality and *bfr* be the blocking factor. Compare the cost function for SELECT operation in the following cases i) when a clustering index is available ii)When a secondary index is available.

ii. Consider a STUDENT file with 20,000 records stored in a disk with fixed length blocks of size 1024 bytes. Each record is of 40 bytes. Assume that in the STUDENT file, there exists a secondary index on key filed, Sid, with X_{SID}=3. There is another file, COURSE_REG, with attributes StudID, CourseID, CourseName and Date of Registration. There are 40,000 records in COURSE_REG file, stored as 4000 blocks. A secondary index on non key key field StudID with X_{STUDID}=4 is available. Let the join selectivity be 1/8 and 6 output records be stored in a block. Find the number of block accesses required for nested join and single loop join for the following query:

$$\textbf{STUDENT} \bowtie_{\text{SID=STUDID}} \textbf{COURSE\_REG}$$

Course Outcome4 (CO4):

1. There are four sites S1, S2, S3 and S4 in a distributed database system with weights 2, 3, 4 and 3 respectively. Assume read quorum value is 6. If a data item x is replicated across these sites and quorum consensus protocol is followed:
   i. find the minimum possible value of write quorum.
   ii. Minimum number of sites locked to perform a read operation

iii. Minimum number of sites locked to perform a write operation

2. Explain 2 phase commit protocol in a distributed environment . What actions would be taken when a site recovers from failure?

Course Outcome 5 (CO5)

1. It is required to represent a University database in XML form. A University has one or more departments. Each department has a name, a specialization, Head of the Department. Several faculty members are working on each department and several courses are run by a department. Each department is uniquely identified by a number(attribute). Name, Area of Specialization each faculty needs to be stored.

   Information such as CourseId, CourseName, Duration and Credits are to be kept about each course. Design a DTD for this University structure

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS017

Course Name: ADVANCED DATABASE

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Given a relation schema R(ABCD) and set of dependencies G={$A \rightarrow B$, $BC \rightarrow D$, $A \rightarrow C$} **(5)**

1. Identify the key.
2. Identify the normal form.
3. Decompose into BCNF.

2. A file has r=36000 STUDENT records of fixed length. Each record has the following fields: Name (25 bytes), SSN (8 bytes), Address (35 bytes), Phone (10 bytes), Date of Birth (8 bytes), Sex(1 bytes), Class code (3 bytes). **(5)**

i. Calculate the record size of R

ii. Calculate the blocking factor bfr and number of file blocks b, assuming an unspanned organization

iii. Calculate the average time it takes to find a record doing a linear search on the file (assume blocks are stored continuously)

iv. Assume a primary index is created with key as SSN and the data pointer needs 8 bytes, find the number blocks required to keep the primary index. Also find the average time required to find a record using the index.

3. Discuss the rules for transformation of query trees and identify when each rule should be applied during optimization **(5)**

4. Explain 2 phase commit protocol in a distributed environment. **(5)**

5. Design an XML document for storing hostel mess food details (meals taken such as breakfast, lunch, dinner) with their charges for the month of June 2022. Charges may vary depending on the food taken. Students can opt not to take any meals on certain days. **(5)**

i Write a sample XML for 2 students for 2 days.

ii Write a XQuery to return the lunch details of all.

iii Create an XSD for the same.

Part B

(Answer any five questions. Each question carries 7 marks)

6. (a) Suppose you are given with a relation schema R(ABCD). Each of the following FDs, assuming they are the dependencies hold over R, state whether or not proposed decomposition of R into smaller relation is a good decomposition. Explain Why? **(5)**

1. AB →B C→A, C→D, decompose into ACD and BC
2. A→ BC, C →AD, A→C, decompose into BCD and AC

(b) What Minimal Cover. Illustrate with an example **(2)**

7. Notown Records has decided to store information about musicians who perform on its albums (as well as other company data) in a database. The company has wisely chosen to hire you as a database designer (at your usual consulting fee of $2500/day). Design a conceptual schema for Notown and draw an ER diagram for your schema. **(7)**

a) Each musician that records at Notown has an SSN, a name, an address, and a phone number.

b) Each instrument used in songs recorded at Notown has a unique identification number, a name and a musical key.

c) Each album recorded on the Notown label has a unique identification number, a title, a copyright date, a format, and an album identifier. Each song recorded at Notown has a title and an author.

d) Each musician may play several instruments, and a given instrument may be played by several musicians.

e) Each album has a number of songs on it, but no song may appear on more than one album.

f) Each song is performed by one or more musicians, and a musician may perform a number of songs.

g) Each album has exactly one musician who acts as its producer. A musician may produce several albums, of course.

8. Consider the following statistics about a relational table, STUDENT(Sid, SName, Branch, CNo). There are 16000 records in 4000 blocks with a blocking factor of 4. There is a secondary index on non key attribute CNo with X_{CNO} = 3. Assume, there are only 100 different courses. We have another relation, COURSES (CId, CName, Credit, Type). There are 100 rows in this table, stored in 20 disk blocks. There exists a primary index on CId with X_{CID} = 1. Assume the selection cardinality for the join attribute is 160. **(7)**

Estimate the cost of join operation (STUDENT $\bowtie_{CNo=CId}$ COURSES) by the following type of join operation (avoid the cost incurred for the storage of resultant records) .

i) nested loop join

ii) single loop join.

iii) nested loop join with a buffer space availability of 12 blocks

9. Consider the bitmap representation of the free-space map, where for each block in the file, two bits are maintained in the bitmap. If the block is between 0 and 30 percent full the bits are 00, between 30 and 60 percent the bits are 01, between 60 and 90 percent the bits are 10, and above 90 percent the bits are 11. Such bitmaps can be kept in memory even for quite large files. **(7)**

i Outline two benefits and one drawback to using two bits for a block, instead of one byte as described earlier in this chapter.

ii Describe how to keep the bitmap up to date on record insertions and deletions.

iii Outline the benefit of the bitmap technique over free lists in searching for free space and in updating free space information.

10. (a) Let r (A, B,C) with tuples {(1, 2, 3), (4, 5, 6), (1, 2, 4), (5, 3, 2), (8, 9,7)} and s(C, D, E) with tuples {(3, 4, 5), (3, 6, 8), (2, 3, 2), (1, 4, 1), (1, 2, 3)} are two relation instances. Compute *r semijoin s* **(4)**

(b) What actions would be taken when a site recovers from failure? **(3)**

11. (a) Assume that a Movie database in XML form is available and title, director, year of release, cost of production as the information stored in it. Let MOVIE, TITLE, DIRECTOR, YEAR, COST are the XML elements, and the element MOVIE has an attribute CATEGORY which indicates the type of movies(*Horror, Comedy, Thriller)*. Similarly, the TITLE has an attribute LAN which indicates the language (*Malayalam, English, Hindi*). A movie can have more than one director. **(4)**

Write XPATH queries for the following

I. List all English Movies

ii. List all movies where language is not specified

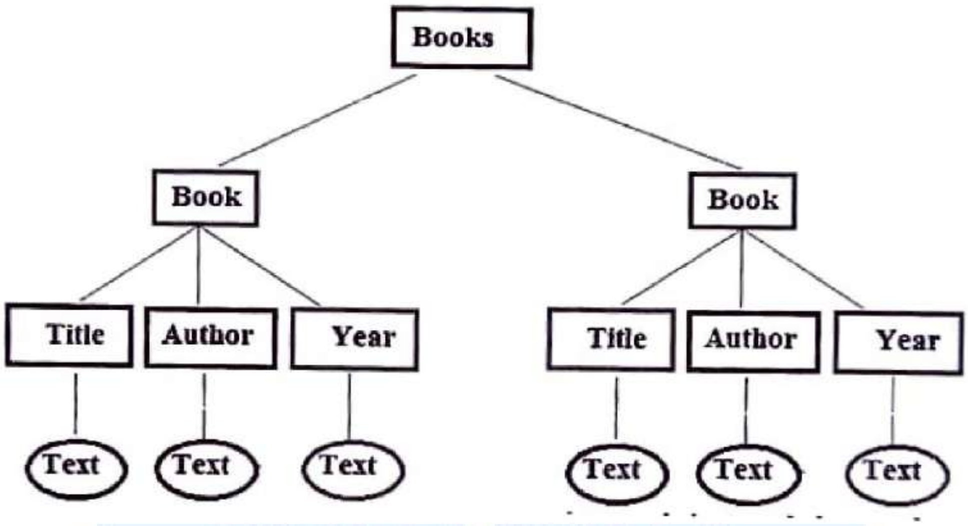
iii. List all movies having two directors

iv. List all *Comedy* type movies in the database

v. List all movies whose cost production is below 10 million.

(b) Explain the terms : i) *Well Formed XML ii) Valid XML* **(3)**

12. Consider the following XML Tree **(7)**



Write an XML schema for the above, and also provide an XQuery expression to get the books published in the year 1992.

Syllabus and Course Plan

221ECS017- ADVANCED DATABASE

Module 1 (Relational Databases – Relational Model, Normalization) 11 Hours

Relational Model Introduction - Structure of Relational Database, database Schema, Keys

The Relational Algebra: Fundamental Operations, The Entity-Relationship model: Entity Set, Relationship Set, Attributes, Constraints: : Mapping cardinalities-E-R Diagrams, Real world Scenarios – ER diagrams. Normalization - The Need for Normalization, Process, Rules for Functional Dependencies, First Normal Form, Second Normal Form, Third Normal Forms, Boyce/Codd Normal Form, Functional Dependencies- Minimal cover, Equivalence, Properties of Relational Decomposition Relational Databases

Module 2: Query Processing and Optimization (8 Hours)

Placing file records on disk- Record types, Record blocking and spanned versus Unspanned records, Hashing techniques –Internal, External hashing for disk files, Indexing and Hashing: Basic concept Ordered Indices, B+ tree Index Files: Structure of a B+- Tree (structure only, algorithms not needed), Indexes on Multiple keys, Hash Indexes, Bitmap indices, Indexing spatial data

Module 3: Introduction to Query Processing and Optimization (6 hours)

Measures of query cost, Algorithms for Selection with cost analysis, Algorithms for Join with cost analysis, Evaluation of expressions, Heuristics in Query Optimization, Optimization of Relational Algebra expressions.

Module 4: Distributed System Architecture (6 Hours)

Introduction to Distributed System architecture, Distributed storage & Distributed file systems

Distributed RDB design & its Transparency, Distributed Transactions, Commit Protocols & Concurrency Control, Distributed Query Processing,

Module 5: XML, XPath, Non-relational Databases ---9Hours

Introduction to Semi-structured Data and XML Databases, XML Data Model – XSD, XML: DTD and XML Schema, XML presentation, XPath Queries , XQuery, Next Generation Databases: Distributed Relational Databases - MongoDB Shading and Replication, Object Relational Systems

Course Plan

No	Topic	No. of Lectures (40 hours)
1	**Module 1 (Relational Databases – Relational Model, Normalization) : 11 hours**	
1.1	Relational Model Introduction - Structure of Relational Database, database Schema, Keys	1
1.2	The Relational Algebra: Fundamental Operations	1
1.3	The Entity-Relationship model: Entity Set, Relationship Set, Attributes, Constraints: : Mapping cardinalities-E-R Diagrams	1
1.4	Real world Scenarios – ER diagrams	1
1.5	Normalization - The Need for Normalization, Process	1
1.6	Rules for Functional Dependencies	1
1.7	First Normal Form, Second Normal Form, Third Normal Forms	1
1.8	Boyce/Codd Normal Form	1
1.9	Functional Dependencies- Minimal cover, Equivalence	1
1.10	Properties of Relational Decomposition	1
1.11	Algorithms for Relational Database Design	1

2	**Module 2: Query Processing and Optimization (8 Hours)**	
2.1	Placing file records on disk- Record types, Record blocking and spanned versus Unspanned records	1
2.2	Hashing techniques –Internal, External hashing for disk files	1
2.3	Indexing and Hashing: Basic concept Ordered Indices	1
2.4	B+ tree Index Files:	1
2.5	Structure of a B+- Tree (structure only, algorithms not needed)	1
2.6	Indexes on Multiple keys	1
2.7	Hash Indexes, Bitmap indices	
2.8	Indexing spatial data	
3	**Module 3: Introduction to Query Processing and Optimization**	**(6 hours)**
3.1	Measures of query cost	1
3.2	Algorithms for Selection with cost analysis	1
3.3	Algorithms for Join with cost analysis	1
3.4	Evaluation of expressions	1
3.5	Heuristics in Query Optimization	1
3.6	Optimization of Relational Algebra expressions	1
4	**Module 4 : Distributed System Architecture (6 Hours)**	
4.1	Introduction to Distributed System architecture	1
4.2	Distributed storage & Distributed file systems	1
4.3	Distributed RDB design & its Transparency	1
4.4	Distributed Transactions	1
4.5	Commit Protocols & Concurrency Control	1

4.6	Distributed Query Processing	1
5	**Module 5: XML, XPath, Non-relational Databases ---9Hours**	
5.1	Introduction to Semi-structured Data and XML Databases	1
5.2	XML Data Model – XSD	1
5.3	XML: DTD and XML Schema, XML presentation	1
5.4	XPath Queries	1
5.5	XQuery	
5.6	Next Generation Databases: Distributed Relational Databases -	1
5.7	Nonrelational Distributed Databases	1
5.8	MongoDB Shading and Replication	1
5.9	Object Relational Systems	1

Reference Books

1. Ramez Elmasri, Shamkant B.Navathe, " Fundamentals of Database Systems ", Pearson Education, 6th Edition, 2007. (Module 1- Chapter 7.1 to 7.7, 14.1 to 14.5, 15.1 to 15.3, Module 2:16.4, 16.8, 17.1 to 17.3, Module : 18.1 to 18.3, 18.7 to 18.9)
2. Abraham Silberschatz, Henry F. Korth, S. Sudarshan," Database System Concepts", McGraw Hill Education, 6th Edition, 2011. (Module 4)
3. Guy Harrison, "Next Generation Databases: NoSQL, NewSQL, and Big Data", Apress, 1st Edition, 14 December 2015.
4. Rob, Peter and Carlos Coronel, "Database Principles: Fundamentals of Design, Implementation and Management", 9th Edition, 2011.
5. Thomas M Connolly and Carolyn E Begg, "Database systems- A Practical Approach to Design, Implementation and Management", Pearson Education, 4th Edition (2014).
6. Ashutosh Kumar Dubay, "Database Management Concepts", S.K. Kataria & Sons, 1st Edition (2012).
7. Raghu Ramakrishnan and Johannes Gehrke, "Database Management Systems", McGraw Hill, 3rd Edition (2014).

221ECS018	CONCEPTS IN CLOUD COMPUTING	CATEGORY	L	T	P	CREDIT
		Program Elective 2	3	0	0	3

Preamble: Cloud computing is the delivery of computing services over the Internet. The syllabus is prepared with a view to equip the students to learn basic concepts in cloud computing - compute, storage, networking. They should gain basic understanding of orchestration, HA and failover. After learning this course computation services can offer faster innovation, flexible resources, and economies of scale.

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

CO 1	Make use of the concepts in cloud computing and OpenStack logical architecture to develop applications (Apply)
CO 2	Explore OpenStack *cloud controller and common services (Apply)*
CO 3	Compare different Open Stack compute service components and storage types (Apply)
CO 4	Analyse *the Open Stack Networking- Connection types and networking services (Analyse)*
CO 5	Analyse the *orchestration, HA and failover in OpenStack (Analyse)*
CO 6	Design, develop, implement and present innovative ideas on cloud computing (Analyse)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	√		√			√	
CO 2	√		√	√	√	√	
CO 3	√		√	√		√	
CO 4	√		√		√	√	
CO 5	√		√			√	
CO 6	√	√	√	√	√	√	√

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	50-80
Analyse	20-40
Evaluate	Can be done through Assignments/projects
Create	Can be done through Assignments/projects

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES: 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS018

Course Name: CONCEPTS IN CLOUD COMPUTING

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Differentiate between private cloud and public cloud? Illustrate the design of OpenStack logical architecture **(5)**

2. *Illustrate* asymmetric clustering and symmetric clustering. *Categorize* the functionalities handled by the cloud controller **(5)**

3. Design the systems of docker containers. **(5)**

4. Derive the procedure for connecting two networks using a virtual router. **(5)**

5. How high-performance, open-source load balancer and reverse proxy for TCP and HTTP applications are obtained. List the HA levels in OpenStack. **(5)**

Part B

(Answer any five questions. Each question carries 7 marks)

6. (a) How the provisioning of VM in OpenStack is organised. Design the system and draw a diagrammatic representation **(4)**

(b) Identify the best practices used in Physical mode design **(3)**

7. Design a method with steps for running OpenStack playbooks **(7)**

8. Explain the deploying swift services. **(7)**

9. Design the architecture of neutron and explain. **(7)**

10. Identify the steps involved in setting a database with high availability **(4)**

11. Compare object storage with NAS/SAN based storage **(7)**

12. Identify the categorization of neutron virtual networks and justify the categories. **(7)**

Syllabus - CONCEPTS IN CLOUD COMPUTING

Module 1: Overview of OpenStack (6 Hours)

Introduction to cloud computing, private cloud, public cloud, hybrid cloud architecture. Cloud Services - Infrastructure as a Service, Platform as a Service, Storage as a Service. Designing OpenStack Cloud Architectural Consideration - OpenStack - The new data center paradigm - OpenStack logical architecture. Nova - Compute service . Neutron - Networking services . Gathering the pieces and building a picture. A sample architecture setup.

Module 2: OpenStack cluster - Controller and common services (5 Hours) OpenStack Cluster – The Cloud Controller and Common Services, Asymmetric clustering, Symmetric clustering. The cloud controller - The keystone service. The nova-conductor service, The nova- scheduler service, The API services, Image management. The network service, The horizon dashboard, The telemetry services.

Module 3: OpenStack compute and Storage (11 Hours) The compute service components-Deciding on the hypervisor OpenStack Magnum project. Segregating the compute cloud Over commitment considerations. Storing instances' alternatives. Understanding instance booting. Planning for service recovery. OpenStack Storage - Block, Object, and File Share-Understanding the storage types. A spotlight on swift. Deploying swift service. Using Block Storage Service Cinder.

Module 4: OpenStack Networking (9 Hours)

The architecture of Neutron. Implementing virtual networks - VLAN, Tunnel based. Virtual Switches, The ML2 Plugin. Neutron Subnets Connecting virtual networks with routers - Configuring the routing service Connecting networks using a virtual router, Connecting to the external world Connectivity from the external world, Associating a floating IP Implementing network security in OpenStack.

Module 5: OpenStack Orchestration, HA and Failover (9 Hours) Orchestration in OpenStack, Heat and its Components. Stacking in OpenStack. OpenStack Orchestration with Terraform. Scope of HA in OpenStack. HA in the database. HA in the Queue, Implementing HA on RabbitMQ

Course Plan

No	Topic	No. of Lectures (40 hrs)
1	**Overview of OpenStack**	**(6 Hours)**
1.1	Introduction to cloud computing, private cloud, public cloud, hybrid cloud architecture.	1
1.2	Cloud Services - Infrastructure as a Service, Platform as a Service, Storage as a Service	1
1.3	Designing OpenStack Cloud Architectural Consideration - OpenStack - The new data center paradigm -OpenStack logical architecture	1
1.4	Nova - Compute service . Neutron - Networking services .	1
1.5	Gathering the pieces and building a picture	1
1.6	A sample architecture setup	1
2	**OpenStack cluster - Controller and common services**	**(5 Hours)**
2.1	OpenStack Cluster – The Cloud Controller and Common Services Asymmetric clustering, Symmetric clustering	1
2.2	The cloud controller - The keystone service. The nova-conductor service	1
2.3	The nova-scheduler service, The API services, Image management.	1
2.4	The network service	1
2.5	The horizon dashboard, The telemetry services	1
3	**OpenStack compute and Storage**	**(11 Hours)**

3.1	The compute service components-Deciding on the hypervisorOpenStack Magnum project.	1
3.2	Segregating the compute cloud	1
3.3	Over commitment considerations.	1
3.4	Storing instances' alternatives. Understanding instance booting.	1
3.5	Planning for service recovery.	1
3.6	OpenStack Storage - -	1
3.7	Block, Object, and File Share	1
3.8	Understanding the storage types.	1
3.9	A spotlight on swift.	1
3.10	Deploying swift service.	1
3.11	Using Block Storage Service Cinder	1
4	**OpenStack Networking**	**(9 Hours)**
4.1	The architecture of Neutron.	1
4.2	Implementing virtual networks - VLAN, Tunnel based.	1
4.3	Virtual Switches, The ML2 Plugin	1
4.4	Neutron Subnets Connecting virtual networks with routers	1
4.5	Configuring the routing service Connecting networks using a virtual router	1
4.6	Connecting to the external world	1
4.7	Connectivity from the external world,	1
4.8	Associating a floating IP Implementing network security in OpenStack- Part 1	1
4.9	Associating a floating IP Implementing network security in OpenStack- Part 2	1
5	**OpenStack Orchestration, HA and Failover**	**(9 Hours)**
5.1	Orchestration in OpenStack,	1
5.2	Heat and its Components.	1

5.3	Stacking in OpenStack	1
5.4	OpenStack Orchestration with Terraform.	1
5.5	Scope of HA	1
5.6	Scope of HA in OpenStack	1
5.7	HA in the database.	1
5.8	HA in the Queue	1
5.9	Implementing HA on RabbitMQ	1

Text Book

1. Omar Khedher, Chandan Datta Chowdhury, Mastering OpenStack, 2nd Edition, Packt Publishing, 2017

Reference Books

1. Tom Fifield, Diane Fleming, Anne Gentle,Lorin Hochstein, Jonathan Proulx, Everett Toews, and Joe Topjian, OpenStack Operations Guide, O'REILY,1/e, 2014.

2. Uchit Vyas, Applied OpenStack Design Patterns, Apress, 1/e, 2016.

3. V. K. Cody Bumgardner, OpenStack in action, Manning, 2016.

4. Amar Kapadia, Sreedhar Varma, Kris Rajana, Implementing Cloud Storage with OpenStack Swift, Packt Publishing, 2014.

5. https://docs.openstack.org/wallaby/?_ga=2.231002015.1428061357.1620834394-

1139122985.1620834394

CODE 221ECS019	STATISTICS FOR DATA SCIENTISTS	CATEGORY	L	T	P	CREDIT
		PEC -2	3	0	0	3

Preamble: This course is intended to systematically master the core concepts in statistics & probability, descriptive statistics, hypothesis testing, regression analysis, analysis of variance, and some advanced regression/machine learning methods such as logistics regressions, polynomial regressions and decision trees. This course helps the students to work with different types of data and implement the techniques and make data-driven decisions

Course Outcomes:

After the completion of the course, the student will be able to

CO 1	Apply the fundamentals of statistics, from bar plots to ANOVAs, regression to k-means, and t-test to non-parametric permutation testing for machine learning, AI, and data science. **(Cognitive knowledge level: Apply)**
CO 2	Visualize the data in different descriptive, inferential, and predictive concepts for relevant stages of data analytics. **(Cognitive knowledge level: Apply)**
CO 3	Analyse the data by making use of concepts such as mean, median, and mode, plus range and IQR and box-and-whisker plots **(Cognitive knowledge level: Apply)**
CO 4	Apply the right statistical technique at appropriate stage of a data analytics project
CO 5	Implement statistical concepts in Python / MATLAB (**Cognitive knowledge level: Apply)**
CO 6	Draw inferences from the data for different machine learning models through hypothesis testing. **(Cognitive knowledge level: Apply)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real-world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources, and state-of-the-art tools to model, analyse and solve practical engineering problems.

PO6: An ability to engage in lifelong learning for the design and development related to the stream-related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓	✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	
CO 5	✓		✓	✓	✓	✓	
CO 6	✓	✓	✓	✓	✓	✓	

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	60-80%
Analyse	20-40%
Evaluate	Attain through Project/Assignments
Create	Attain through Project/Assignments

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

The evaluation shall only be based on application, analysis, or design-based questions (for both internal end-semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer-reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include a minimum of 80% of the syllabus.

Course-based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation, and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts: Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem-solving and quantitative evaluation), with a minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

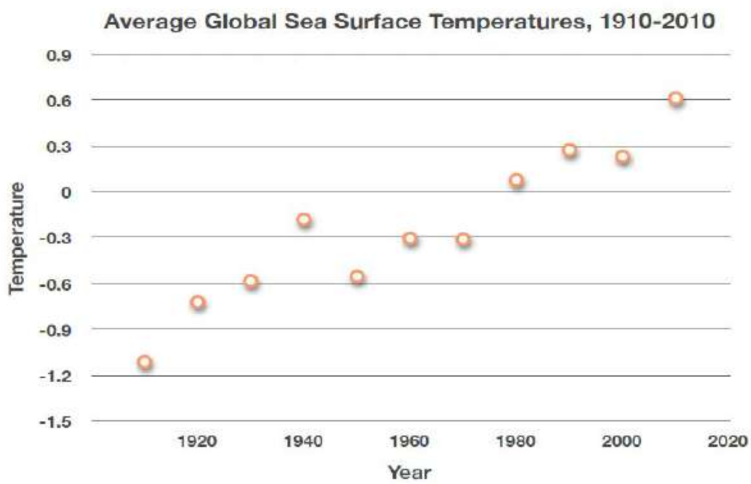
Course Level Assessment Questions

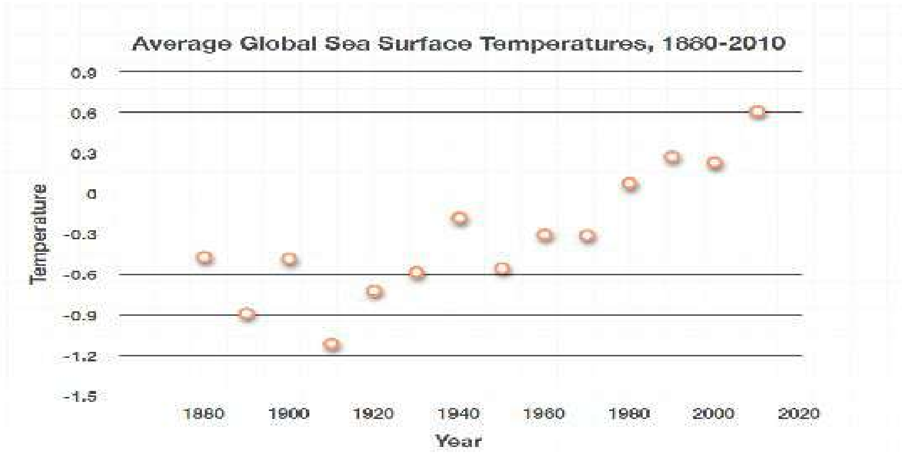
Course Outcome 1 (CO1):

1. Identify the variables in the following data description and classify the variables as categorical or quantitative. If the variable is quantitative, list the units.

   "The Indianapolis 500 is a car race that's been taking place since 1911 and is often scheduled to take place over Memorial Day weekend. The race takes place at the Indianapolis Motor Speedway and a driver needs to complete 200 laps that cover a distance of 500 miles. Race results are reported by driver number, the driver's name, the type of car the driver uses, and the time to the nearest ten-thousandth of a second. If a driver doesn't finish the race, their number of laps completed is recorded instead of the time to complete the race."

2. Compare the scatterplots. The second graph includes extra data starting in 1880. How does this compare to the plot that only shows 1910 to 2010? Explain trends in the data, and how the regression line changes by adding in these extra points. Which trend line would be best for predicting the temperature in 2050?





3. Calculate and interpret the correlation coefficient for the data set

x	y
54	0.162
57	0.127
62	0.864
77	0.895
81	0.943
93	1.206

Course Outcome 2 (CO2)

1. Explain Descriptive analytics, Diagnostic analytics, Predictive Analytics, and Prescriptive analytics
2. Sarah's monthly sales to date are shown in the ogive. Signify the meaning of circled point

3. Bethany started a sit-up program so that she can do 200 sit-ups in a day. At the end of week 6 she'll have completed 1,685 sit-ups. Create an ogive of the data.

Week	Number of sit-ups
Week 1	350
Week 2	455
Week 3	600
Week 4	540
Week 5	1,275
Week 6	1,685

Course Outcome 3(CO3):

1. a. Elaborate Normal distributions and z-scores
   b. Let's say the mean finishing time for male speed skaters in the winter Olympics on the 500-meter track is 70.42 seconds, with a standard deviation of 0.34 seconds (the data is normally distributed). What is the maximum time a skater can post if he wants to skate faster than 95% of his competitors?
2. a. Illustrate the advantages of the bar chart over the pie chart
   b. List the charts for categorical variables and quantitative variables and Illustrate

3. Create the total-relative frequency table for the data, and then answer the question: Carl is in charge of creating an activity for the students in his college dorm. If Carl wants the highest possible turnout, which activity should he choose? Why?

	Movie	Bowling	Pizza Party
Male	20	40	55
Female	35	50	62

Course Outcome 4 (CO4):

1. Explain the techniques you will use for your data analysis before you collect any data

2. The first stage in any analysis should be to describe your data and the population from which it is drawn. The statistics appropriate for this activity fall into three broad groups and depend on the type of data you have.

Answer the following

a. What do you want to have to look at the distribution, to describe the central tendency, and describe the spread

b. With what type of data

c. Appropriate techniques

3. Explain the appropriate Statistical techniques to find the Differences between groups and variables and relationship between the variables

Course Outcome 5 (CO5):

1. Explain the Central Limit Theorem and Implement it in Python / MATLAB

2. Explain KNN and PCA and Implement them in Python / MATLAB

Course Outcome 6 (CO5):

1. Explain the significance of hypothesis testing for machine learning and the parameters of hypothesis testing

2. A fast-food restaurant is implementing new workplace policies with the goal of increasing employee satisfaction by 2 points on a scale of 1 to 10. The restaurant surveys 10 employees, asking them both before and after the policies are enacted to rate their workplace satisfaction on the 1 – 10 scale and records the results in the table below.

Employee	1	2	3	4	5	6	7	8	9	10
Before x_1	3	3	5	7	1	0	2	6	6	5
After x_2	3	6	9	7	3	5	5	5	9	9
Difference, d	0	3	4	0	2	5	3	-1	3	4
d^2	0	9	16	0	4	25	9	1	9	16

Can the restaurant say at 5% significance that the policies increased employee satisfaction by 2 points?

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS019

Course Name: STATISTICS FOR DATA SCIENTISTS

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. A company is analyzing the results from a recent survey about why people left their employment. The results are shown in the data table below. In general, is a bar graph or a pie chart a better choice to display the data? Why? **5**

Reasons for leaving a job	
Reduced job duties	30%
Company restructuring	15%
Too much travel time	12%
Looking for more opportunity	11%
Need more personal time	9%
Poor expected company growth	8%
The job was a contract or short term	8%
Need more of a challenge	5%
Other	2%

2. John stops at the local gas station and decides to buy lottery tickets. Each ticket has a 20% chance of being a winner. He will buy a lottery ticket and check to see if it's a winner. He'll collect his money and be done if it's a winner. If it's not a winner, he'll buy another. He'll repeat this until he gets a winning ticket. But if he hasn't won by his fifth ticket, he won't buy any more tickets. Let L be the number of lottery tickets John will buy, then find E(L). **5**

3. Estimate the effect of bias on sampling. **5**

4. Compare correlation and regression **5**

5. Explain Signal detection theory **(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6. (a) Appraise the happenings to our measures of central tendency and spread when we make changes to our data set(mean, median, mode, range, and IQR) **(3)**

i. Changing the entire data set
ii. Adding or removing a data point from the set

(b) Illustrate the appropriate visualization technique to represent a data set when we want to show the median and spread of the data at the same time **(2)**

(c) Describe the importance of Box-and-whisker plots **(2)**

7. (a) Illustrate measures of central tendency **(3.5)**

(b) Illustrate the measures of spread (Range, interquartile range (IQR), variance, standard deviation) **(3.5)**

8. (a) Two factories A and B produce heaters for car seats. A customer received a defective car seat heater and the manager at factory B would like to know if it came from her factory. Use the table below to determine the probability that the heater came from factory B. **(2.5)**

Factory	% of production	Probability of defective heaters
A	0.55	0.020 $P(D \mid A)$
B	0.45	0.014 $P(D \mid B)$

(b) Explain Poisson process **(2)**

(c) There are 30 students in a Kindergarten class and each one of them has a 4% chance of forgetting their lunch on any given day. What is the probability that exactly 5 of them will forget their lunch today? **(2.5)**

9. (a) You are planning a day at the beach, but the morning is cloudy. 64% of all rainy days start off cloudy, but cloudy mornings are common (55% of days start cloudy). This month is usually a dry month and only 18% of the days tend to be rainy. What's the chance that it will rain during your day at the beach? **(3)**

(b) The time it takes students to complete multiple choice questions on an AP Statistics Exam has a mean of 55 seconds with a standard deviation of 12 seconds. If the exam consists of 40 multiple choice questions, find the mean total time to finish the exam. Then find the standard deviation in the total time. What assumption must be made? **(2)**

(c) We toss a fair coin 15 times and record the number of tails. Is this experiment modeled by a binomial random variable? If it isn't, explain why.If it is, determine its parameters n and p and express the binomial random variable as $X \sim B(n, p)$. **(2)**

10. (a) Suppose we want to determine whether the mean height of men is significantly higher than the mean height of women in a certain city, so we randomly sample 100 men and 100 women. Given the mean and standard deviation of both samples below, use the critical value approach to say whether men are significantly taller than women at a 1% level of significance. **(7)**

Men	Women
$n_1 = 100$	$n_2 = 100$
$\bar{x}_1 = 69.5$ inches	$\bar{x}_2 = 67.8$ inches
$s_1 = 1.25$ inches	$s_2 = 1.12$ inches

(3)

11. (a) Three types of batteries were tested for battery life. See the battery lives in the table below. In constructing the ANOVA table, what will be the values of factor and error degrees of freedom? **(3.5)**

C1	C2	C3
Battery 1	Battery 2	Battery 3
90	102	102
85	98	82
92	97	85
95	103	75
88	107	92

(b) Explain confidence interval and its imporance **(5)**

12. (a) Explain the statistical power and its importance **(2)**

(b) Use the Average Global Sea Surface Temperatures data shown in the table to create a line of best fit for the data. Consider 1910 as year 10. Use the equation to predict the average global sea surface temperature in the year 2050. **(5)**

Year	Temperature, F
1910	-1.11277
1920	-0.71965
1930	-0.58358
1940	-0.17977
1950	-0.55318
1960	-0.30358
1970	-0.30863
1980	0.077197
1990	0.274842
2000	0.232502
2010	0.612718

Syllabus

Mod	Contents	hrs
I	**DESCRIPTIVE STATISTICS:** DATA, Types of data, data, Sample vs. population data, Samples (N=1 & N>1 studies), **Visualizing Data**: Bar plots. Box-and-whisker plots, Boxplots of normal and uniform noise, Histograms, Histogram proportion, Pie charts, Implementation, descriptive vs. inferential statistics, Accuracy, precision, and resolution. Data distributions, histograms of distributions, Measures of central tendency, central tendencies with outliers, Measures of dispersion Interquartile range, QQ plots, Statistical "moments", Histograms: Violin plots, Shannon entropy, entropy, and number of bins, Implementation in Python	7
II	**DATA NORMALISATION AND PROBABILITY:** Z-score standardization, Min-max scaling, Removing outliers: z-score method, modified z-score method, z vs. modified-z, Multivariate outlier detection, Euclidean distance for outlier removal, Removing outliers by data trimming, Non-parametric solutions to outliers, Implementation **Probability:** Computing probabilities, Probability mass vs. density, PDF, CDF, creating sample estimate distributions, Monte Carlo sampling, Sampling variability, noise, and other annoyances, Expected value, Conditional probability, and Tree diagrams, The Law of Large Numbers, The Central Limit Theorem,, Implementation **Random Variables:** Discrete RV, Binomial Poisson, Bernoulli, and Geometric random variables	8
III	**SAMPLING & HYPOTHESIS TESTING: Sampling:** Types of studies, Sampling, and bias, Sampling distribution of the sample mean, Conditions for inference with the SDSM, Sampling distribution of the sample proportion, Conditions for inference with the SDSP, **Hypothesis Testing:** Independent and Dependent Variables, models, residuals, Sample distributions under null and alternative hypotheses, P-values: definition, tails, and misinterpretations, P-z combinations that you should memorize, Degrees of freedom, Type 1 and Type 2 errors, Parametric vs. non-parametric tests, Multiple comparisons and Bonferroni correction, Statistical vs. theoretical vs. clinical significance, Cross-validation, Statistical significance vs. classification accuracy, **The T-Test Family:** Purpose and interpretation, One-sample t-test, The role of variance, Two-samples t-test, Importance of N for t-test, Wilcoxon signed-rank (nonparametric t-test), Mann-Whitney U test (nonparametric t-test), U test, Permutation testing for t-test, significance, Python Implementation	8
IV	**CONFIDENCE INTERVALS & ANOVA: Confidence Intervals on Parameters:** Computing confidence intervals via formula, Confidence intervals via bootstrapping (resampling), Misconceptions about confidence intervals. CORRELATION: Motivation and description of correlation, Covariance, and correlation: formulas, Correlation matrix, correlation to the covariance matrix, Partial correlation, The problem with Pearson, Nonparametric correlation: Spearman rank, Fisher-Z transformation for correlations, Spearman correlation, the confidence interval on the correlation, Kendall's correlation for ordinal data, The subgroups correlation paradox, Cosine similarity, **Analysis Of Variance:** Sum of squares, The F-test and the ANOVA table, The omnibus F-test and posthoc comparisons, The two-way ANOVA, One-way ANOVA example, Two-way ANOVA example, **Regression:** Introduction to GLM / regression, Least-squares solution	8

	to the GLM, Evaluating regression models: R2 and F, Simple regression, implementation	
V	**REGRESSION, CLUSTERING, AND PCA: Regression:** Multiple regression, Standardizing regression coefficients, Polynomial regression models, Logistic regression, Under and over-fitting, comparing "nested" models, missing data, **Statistical Power, And Sample Tests:** Importance of statistical power, Estimating statistical power and sample size, Compute power and sample size using G*Power, **Clustering And Dimension-Reduction-**K-means clustering, K-means, and normalization, K-means on a Gauss blur, Clustering via dbscan, dbscan vs. k-means, K-nearest neighbor classification, Principal components analysis, K-means on PC data, independent components analysis, **Signal Detection Theory:** The two perspectives of the world, d-prime, Response bias, F-score, Receiver operating characteristics (ROC), Python Implementation	9

Course Plan

S.NO	TOPIC	NO. OF LECTURES (40 hrs)
	Module 1- Descriptive Statistics - 7 Hours	
1.1	Types of data, Sample vs. population data, Samples,	1
1.2	**Visualizing Data**: Bar plots. Box-and-whisker plots, Boxplots of normal and uniform noise, Histograms,	1
1.3	Histogram proportion, Pie charts, implementation	1
1.4	descriptive vs. inferential statistics, Accuracy, precision, and resolution. Data distributions, histograms of distributions,	1
1.5	Measures of central tendency, central tendencies with outliers	1
1.6	Measures of dispersion Interquartile range, QQ plots, Statistical "moments	1
1.7	Histograms: Violin plots, Shannon entropy, entropy, and number of bins, code	1
	Module 2- Data Normalisation and Probability- 8 Hrs	
2.1	**DATA NORMALISATION AND PROBABILITY:** Z-score standardization, min-max scaling, removing outliers: z-score method, The modified z-score method, z vs. modified-z,	1
2.2	Multivariate outlier detection, Euclidean distance for outlier removal, Removing outliers by data trimming, non-parametric solutions to outliers,	1
2.3	**Probability:** Computing probabilities,	1
2.4	Probability mass vs. density, PDF, CDF. CDFs for various distributions,	1
2.5	Creating sample estimate distributions, Monte Carlo sampling, Sampling variability, noise, and other annoyances, Expected value, Conditional probability, and Tree diagrams,	1
2.6	The Law of Large Numbers, The Central Limit Theorem, Implementation	1
2.7	**Random Variables:** Binomial and Poisson RV,	1

2.8	Bernoulli, and Geometric random variables	1
	Module 3-Sampling and Hypothesis Testing-7 Hrs	
3.1	**Sampling:** Types of studies, Sampling, and bias, Sampling distribution of the sample mean,	1
3.2	Conditions for inference with the SDSM Sampling distribution of the sample proportion, Conditions for inference with the SDSP,	1
3.3	**Hypothesis Testing:** Independent and Dependent Variables, models, residuals, Sample distributions under null and alternative hypotheses,	1
3.4	P-values: definition, tails, and misinterpretations, P-z combinations that you should memorize, Degrees of freedom, Type 1 and Type 2 errors,	1
3.5	Parametric vs. non-parametric tests, Multiple comparisons, and Bonferroni correction, Statistical vs. theoretical vs. clinical significance, Cross-validation, Statistical significance vs. classification accuracy,	1
3.6	**The T-Test Family:** Purpose and interpretation, One-sample t-test, The role of variance, Two-samples t-test,	1
3.7	Importance of N for t-test, Wilcoxon signed-rank (nonparametric t-test),	1
3.8	Mann-Whitney U test (nonparametric t-test), U test, Permutation testing for t-test, significance, Python Implementation	1
	Module 4-Confidence Intervals, Correlation and Anova- 8 Hrs	
4.1	Confidence Intervals on Parameters: Computing confidence intervals via formula, Confidence intervals via bootstrapping (resampling), Misconceptions about confidence intervals	1
4.2	**Correlation:** Motivation and description of correlation, Covariance, and correlation: formulas, Correlation matrix,	1
4.3	correlation to the covariance matrix, Partial correlation, The problem with Pearson, Nonparametric correlation: Spearman rank, Fisher-Z transformation for correlations,	1
4.4	Spearman correlation, the confidence interval on the correlation, Kendall's correlation for ordinal data, The subgroups correlation paradox, Cosine similarity,	1
4.5	**Analysis of Variance:** Sum of squares, The F-test and the ANOVA table, The omnibus F-test and posthoc comparisons,	1
4.6	The two-way ANOVA, One-way ANOVA example,	1
4.7	Two-way ANOVA example,	1
4.8	**Regression:** Introduction to GLM / regression, Least-squares solution to the GLM, Evaluating regression models: R2 and F, Simple regression, code	1
	Module 5- Regression, Clustering and PCA-9 Hrs	
5.1	**Regression:** Multiple regression, Standardizing regression coefficients	1
5.2	Polynomial regression models, Logistic regression,	1
5.3	Under and over-fitting, comparing "nested" models, missing data,	1
5.4	**Statistical Power and Sample Tests** Estimating statistical power and sample size, Compute power and sample size using G*Power,	1

5.5	**Clustering And Dimension-Reduction-**K-means clustering, K-means, and normalization, K-means on a Gauss blur, Clustering via dbscan	1
5.6	dbscan vs. k-means, K-nearest neighbor classification, Principal components analysis,	1
5.7	K-means on PC data, independent components analysis,	1
5.8	**Signal Detection Theory:** d-prime, Response bias	1
5.9	F-score, Receiver operating characteristics (ROC), Python Implementation	1

TEXTBOOKS:

1. A. Abebe, J. Daniels, J. W. Makean, "Statistics and Data Analytics", Statistical Computation Lab, Western Michigan University, Kalamazoo.2001
2. Peter Goos and David Meintrup, 'Statistics with JMP: Graphs, Descriptive Statistics, and Probability, WILEY 2015
3. Peter Goos and David Meintrup, 'Statistics with JMP: Hypothesis Tests, Anova, and Regression' WILEY 2016
4. Bruce Ratner, 'Statistical and Machine-Learning Data Mining-Techniques for Better Predictive Modeling and Analysis of Big Data, Third Edition, CRC Press, Tailor and Francis group, 2017
5. Charles Wheelan, 'Naked Statistics_ Stripping the Dread from the Data', W.W. Norton Company, New York, 2014

REFERENCES:

1. Jim Albert, "Bayesian Computation with R", 2nd Edition, Springer 2009
2. Trevor Hasti, Robert Tibshirani, Jerome Friedman, "Data Mining, Inference and Statistics", 2nd Edition, Springer Series in Statistics 2008

CODE 221ECS020	ETHICS FOR DATA SCIENTISTS	CATEGORY	L	T	P	CREDIT
		PROGRAM ELECTIVE 2	3	0	0	3

Preamble:

This course is intended to provide an introduction to critical and ethical issues using data and its implications in the society. This course helps the learners to understand the benefits and drawbacks of using data while using them for making predictions by understanding the structure of ethics, law, and societal values. Also, this course blends social and historical perspectives on data with ethics, policy, and case examples to help students develop a workable understanding of current ethical issues in data science.

Course Outcomes:

After the completion of the course, the student will be able to

CO 1	Applying the concept of ethics, and the necessity and the benefit of adopting shared ethical principles in Data Science (**Cognitive knowledge level: Apply)**
CO 2	Analyze the role of an IRB in a human subject's study and the difference between data collected for business purposes versus research purposes and relate how this changes the requirement for IRB approval. (**Cognitive knowledge level: Apply)**
CO 3	Distinguish between the three main categories of intellectual property and identify the data owner (**Cognitive knowledge level: Apply)**
CO 4	Describe the reasonable expectation of privacy relates to data collection and recognize the voluntary limits on the use of data that arise out of social consensus. (**Cognitive knowledge level: Apply)**
CO 5	Apply data ethics while practicing Data Science. (**Cognitive knowledge level: Apply)**
CO 6	Identify invalid data, incorrect models, bias in algorithms, and bad analysis conducted on good data (**Cognitive knowledge level: Apply)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1			✓			✓	
CO 2			✓			✓	
CO 3			✓			✓	
CO 4			✓			✓	

CO 5			✓			✓	✓
CO 6			✓			✓	

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	50-80%
Analyse	20-40%
Evaluate	Assignments/Projects
Create	Assignments/Projects

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer-reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation: 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include a minimum 80% of the syllabus.

Course-based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation, and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts: Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem-solving and quantitative evaluation), with a minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions
Course Outcome 1 (CO1):

1. Illustrate an example of a situation, in which we all, as a society, are better off, because we agree to behave ethically.
2. Company X has learned about Facebook's mood manipulation experiment and believes that a happy person is much more likely to buy than a grumpy one. Therefore, it has designed its website to tell heart-warming stories in callout boxes on every page. These stories, at best, are tangentially related to the products being sold on the page. They A/B test this website before launch to see if the story boxes do have the intended effect. They find that the boxes do have the desired effect of increasing sales. They then adopt the new website design with the story boxes, and they write an article describing their findings in a Marketing Journal.

Discuss:

Does Company X need to inform its customers about this effort? To what extent? Does it need to obtain consent? If so, for what? If you answered YES to the consent question above, what is the smallest change to the scenario described above that would make you change your answer to NO?

3. You go to the bus stop, and everyone is patiently in line waiting for the bus. Rather than wait in line, you just jump onto the bus when it arrives. Discuss the ethics about your behavior- whether it is legal/ethical/ unethical/ unethical and legal/ ethical and illegal etc

4. You conduct research on user interface design. You wish to evaluate a new layout you have developed for presenting the results of a web search. For this purpose, you need to get the opinions of several users. Even though the users of your new interface have no possibility of suffering any harm, and furthermore your test is no more intrusive than the A/B testing performed by so many web companies, Discuss- is it true that you are nevertheless required to obtain IRB clearance?

Course Outcome 2 (CO2)

1. Creative Commons has a set of standard copyright licenses that are used widely. This course as a whole is released CC-BY-NC, which means it can be reproduced with attribution (BY) for non-commercial use (NC). Individual components are released CC BY-NC-ND, which means they can be reproduced with attribution (BY) for non-commercial use (NC) without making any changes (ND = no derivatives). Discuss whether it is OK to reuse, with attribution, a single video from this course in your own (non-commercial) presentation.

2. You conduct research on user interface design. You wish to evaluate a new layout you have developed for presenting the results of a web search. For this purpose, you need to get the opinions of several users. Even though the users of your new interface have no possibility of suffering any harm, and furthermore your test is no more intrusive than the A/B testing performed by so many web companies, is it true that you are nevertheless required to obtain IRB clearance?

3. Analyse- You have designed a human subjects experiment with appropriate provision for informed consent, and you submit this to the appropriate IRB. The IRB will automatically

approve your experiment since you have demonstrated you will properly obtain informed consent.

Course Outcome 3(CO3):

1. If teenagers knew that their parents could have access to all their social media posts, then the teenagers would likely be very careful about what they post - Asses

2. I agree to pose for some photographs you take with the promise that you will keep these photos private. Some years later, you change your mind and publish these photos. Since you own these photos, are you within your rights?, Evaluate the action.

Course Outcome 4 (CO4):

In terms of undesired use of data, there are three distinct steps. Justify your answer for each of the following three questions about whether it violates privacy.

1. Undesired collection of personal data violates privacy.

2. Undesired analysis of previously collected personal data violates privacy.

3. Undesired dissemination of previously collected personal data violates privacy.

4. A major shortcoming of all-or-nothing data access policies (e.g., when an app wants access to your location data) is that it defeats privacy because you have no control over any data you choose to share with the app- Discuss

Course Outcome 5 (CO5):

1. It turns out that the government funding for public health departments is computed on a formula that is heavily dependent on the number of cases of flu. For efficiency, the government decides to adopt Google flu numbers for this parameter. If you run a public health department, you seek to maximize your funding by asking the public in your county to perform searches for flu. Will this work? Evaluate.

2. The university in the preceding question conducts some additional investigation, and determines that both the mean and the median scores obtained by minority applicants on the standardized test are substantially lower than the corresponding mean and median for other students. Based on this fact, in conjunction with the facts from the preceding question, can we conclude that the test is unfair to a minority applicant? Justfy

3. A university uses performance on a standardized test as the only scoring mechanism used to admit applicants. The university observes that it is admitting far fewer minority students than their proportion in the population at large. Based on only these facts, can we conclude that the test is unfair? Justify your answer

Course Outcome 6 (CO6):

1. Your city has decided to make property tax payment data semi-public: you just have to enter your property identifier to get that information.

Your neighbor has a small business that you have invested in, and are feeling nervous about. You enter your neighbor's property ID into the city system to check on your neighbor's tax payments. You find that he has missed paying the last two quarters, after many years of paying on time. You suspect a cash flow crunch in his business and ask for your loan back.

Your neighbor is forced to sell some business assets to pay you back. Another investor sees this sale of business assets, and also decided to liquidate her investment. In this manner, problems snowball, until your neighbor is driven out of business.

Whose fault is it, if any? Identify specific steps where some ethical rule was violated.

2. You work for a major cell phone service provider and have access to large volumes of detailed location data for your customers. One day, you are able to correlate location with building footprints and hence determine whether the cell phone user (your customer) is indoors or outdoors. On this basis, you obtain analytical results that lead to a new signal amplification algorithm that is amazingly effective in improving call quality. You surprise yourself, your boss, and your company, with these results. Does this analysis violate the "Do Not Surprise" rule? Evaluate

3. Seeking to expand their business and improve their product, suppose that Amazon sends a survey to all Kindle owners asking them what they like and dislike about their Kindle. What validity concerns would you have about the survey results obtained? If the primary goal is to grow Kindle sales, what could Amazon do to get more valid data?

Model Question Paper

QP CODE:		
Reg No: _______________		
Name: _________________		**PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M. TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS020

Course Name: ETHICS FOR DATA SCIENTISTS

Max. Marks : 60 Duration: 2.5 Hours

PART A

Answer All Questions. Each Question Carries 5 Marks

1.	Briefly explain the importance of Institutional Review Boards and Independent Ethics Committees	**5**
2.	Discuss the various types of Intellectual Property Rights	**5**

3.	a. I am a fan of art called mauve pottery. I painstakingly create a directory of mauve potters around the world. I publish this directory, copyrighted, with all rights reserved. Is it OK for you to make a copy of this directory for use in your (non-commercial) class? b. In the example of the preceding question, is it OK for you to publish a new version of this directory showing the locations of all mauve potters on a map? c. Many psychology experiments are conducted on the university campus by academic researchers. The human subjects recruited tend to be college students, who are generally younger and smarter than the population as a whole. This is clearly not a representative sample of the general population. To show the universal validity of an important effect, researchers went off-campus to a nearby city and recruited volunteer subjects by offering a small cash incentive for their time. i. The second experiment is a good random sample of the population ii. The second experiment is not a good random sample either, and so is pointless. iii. The second experiment is not a good random sample but is still valuable. Explain and justify your answer.	**5**
4.	a. In machine learning, it is common practice to use k-fold cross-validation, where the data set is divided into k parts. k-1 of these are used for training and the remaining part is used for testing. And this can be repeated k times, leaving out a different part each time. Appraise is this a good way to measure how well the learned model predicts the test data (labels, values, or whatever is being predicted)? b. Based on face recognition technology applied to records from in-store video cameras, Fancy Store is immediately able to identify you when you enter their store if you have shopped there before. If you are identified as a high-value shopper, from your previous purchasing history in the store, a personal shop assistant is immediately assigned to stay with you during your visit and help you choose and locate the items you want. Ordinary shoppers, not identified as high-value shoppers, do not get the same service. Is this unfair? Why or why not?	**5**

5.	a. A travel website has empirically determined that Mac users are more willing to pay for higher-priced hotel rooms. Therefore, the website modifies the default order in which hotels are shown, with higher-priced hotels ranking slightly higher for Mac users than for other PC users. Analyze: Is this reordering “discrimination” ethical? b. A leading algorithm-based employment agency determines, based on data analysis, that candidates with straight hair make more reliable employees than candidates with curly hair. They use this as a criterion (one among many, but with significant weight to this one) in choosing which candidates to interview, using submitted photographs with the application as their basis to determine whether hair is straight or curly. They do not tell prospective candidates what criteria they are using. This is unethical.		**5** **(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	(a)	Analyze the impact of not following the data ethics in business	**(7)**
7.	(a)	Discuss the importance of ethics-centricity in data projects	**(7)**
8.	(a)	Appraise Modern Privacy Risks and Protection Strategies in Data Analytics	**(7)**
9.	(a)	Discuss the case study of targeted advertisement, whether it is helpful or annoying. If annoying, is there any method digitally to avoid that?	**(3)**
	(b)	Discuss the Case Study of data privacy breach: Sneaky Mobile Apps	**(4)**
10.	(a)	a. Explain choosing attributes that we will use to achieve our objectives and methods of measuring them.	**(4)**
	(b)	b. Explain the errors in the data processing	**(3)**
11.	(a)	aExplain the need for algorithm fairness	**(2)**
	(b)	Discuss removing bias from hiring algorithms	**(5)**

12.	(a)	Discuss the societal consequences of Data Science that we should be concerned about even if there are no issues with fairness, validity, anonymity, privacy, ownership or human subjects research	**(3.5)**
	(b)	Discuss Social credit Scores and its societal consequences	**(3.5)**

Syllabus

MODULE NO	CONTENT	HOURS
I	**Introduction:** Ethics, Definition, what is Data Ethics? Need of Data Science Ethics, Examples, Case Study and Discussion. Human Subject Research and Informed Consent, Cases, US Institutional Review Board (IRB), IRB in India, Limitations of Informed Consent, Case Study and Discussion.	8
II	**Data Ownership and Privacy:** Data ownership, Limits of Recording Data and Using Data, Intellectual Property rights, Privacy: Introduction, History, Degrees, Privacy Risks, Case Studies- Targeted Ads, The Naked Mile. Sneaky Mobile Apps, and Discussion, Anonymity: De-Identification, Case Studies, and Discussion	8
III	**Data Validity:** Validity: Introduction, Choices of Attribute and Pressures, Errors in Data Processing, Errors in Model Design, Managing Change, Case Study- Three Blind Mice, Algorithms and Race, Algorithms in the Office, GermanWings Crash, Google Flu, and Discussion	8
IV	**Algorithmic Fairness:** Algorithm Fairness: Introduction, Correct and Misleading results, Hiring algorithms, P Hacking, Case Study- High Throughput Biology, Geopricing, Your Safety Is My Lost Income and Discussion- fairness of applying face recognition to give preference to the often visiting customers in business	8
V	**Societal Consequences:** Introduction, Societal Impact, Ossification, Surveillance, Code of ethics, Wrap Up, Case Studies- Social Credit Scores, Predictive Policing, and Discussion	8

Course Plan

No	Topic	No. of Lectures ()
Module 1 (Introduction to Ethics)- 8 hours		
1.1	Ethics, Definition, what is Data Ethics?	1
1.2	Need of Data Science Ethics.	1
1.3	Examples, Case Study and Discussion	1
1.4	Human Subject Research and Informed Consent, Cases, US Institutional Review Board (IRB), IRB in India,	1
1.5	Limitations of Informed Consent,	1
1.6	Case Study and Discussion	1
1.7	Case Study and Discussion	1
1.8	Case Study and Discussion	1
Module 2 (Data Ownership and Privacy)- 8 hours		
2.1	Data Ownership and Privacy	1
2.2	Data Ownership and Privacy	1
2.3	Intellectual Property rights,	1
2.4	Privacy: Introduction, History, Degrees.	1
2.5	Privacy Risks, Case Studies, and Discussion,	1
2.6	Targeted Ads, The Naked Mile. Sneaky Mobile Apps	1
2.7	Anonymity: Introduction, De-Identification,	1

2.8	Case Studies and Discussion	1
Module 3 (Data Validity)- 8 hours		
3.1	Validity: Introduction, Choices of Attribute and Pressures,	1
3.2	Errors in Data Processing,	1
3.3	Errors in Model Design,	1
3.4	Managing Change	1
3.5	Managing Change	1
3.6	Case Study- Three Blind Mice, Algorithms and Race	1
3.7	Case Study- Algorithms in the Office	1
3.8	Case Study- GermanWings Crash, Google Flu, and Discussion	1
Module 4 (Algorithmic Fairness)- 8 hours		
4.1	Algorithm Fairness: Introduction,	1
4.2	Correct and Misleading results.	1
4.3	Correct and Misleading results.	1
4.4	P Hacking,	1
4.5	P Hacking,	1
4.6	Case Study- High Throughput Biology, Geopricing	1
4.7	Case Study- Your Safety Is My Lost Income	1
4.8	Case Study- Discussion- fairness of applying face recognition to give preference to the often-visiting customers in business	1

	Module 5 (Societal Consequences)	
5.1	Introduction,	1
5.2	Societal Impact,	1
5.3	Ossification,	1
5.4	Surveillance	1
5.5	Code of ethics,	1
5.6	Code of ethics, Wrap Up,	1
5.7	Case Studies- Social Credit Scores, and Discussion	1
5.8	Case Studies- Predictive Policing, and Discussion	1

Reference Books

Books:

1. DJ Patil, Hilary Mason, Mike Loukides, 'Ethics and Data Science,', O'Reilly Media, Inc.

2. Shannon Vallor, Ph.D. William J. Rewak, S.J. Professor of Philosophy, 'An Introduction to Data Ethics' Santa Clara University

3. Kord Davis 'Ethics of big data' ISBN: 9789350238806, 9350238802

4. Samiksha Shukla, Jossy P. George, Kapil Tiwari, Joseph Varghese Kureethara, 'Data Ethics and Challenges' Springer, ISBN: 978-981-19-0752-4

5. John Havens 'Artificial Intelligence: Embracing Our Humanity to Maximize Machines' TarcherPerigee. ISBN-10 : 0399171711

URLs

MOOCS: https://courses.edx.org/courses/coursev1:

MichiganX+DS101x+1T2018/courseware/94ac457869964552a69a3f37ba579954/67
1f46458 36145eea658edbf9298be64/

221ECS021	SPEECH PROCESSING	CATEGORY	L	T	P	CREDIT
		PROGRAM ELECTIVE-2	3	0	0	3

Preamble

The course aims to introduce the student to the various aspects of speech processing including modelling of human speech. The topics covered in the course includes Computational Phonology, Models of Spelling and Pronunciation, speech synthesis and speech recognition. It helps the learners to develop application involving speech processing.

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

CO 1	Analyse the different aspects of production of speech in humans (**Cognitive Knowledge Level: Analyse)**
CO 2	Make of use the methods used for spelling error detection and correction (**Cognitive Knowledge Level: Apply**)
CO 3	Illustrate the various models for speech pronunciation variations (**Cognitive Knowledge Level: Apply**)
CO 4	Make use of the different models for recognizing human speech and converting into equivalent text (**Cognitive Knowledge Level: Apply**)
CO 5	Comprehend the processes involved in the acoustic processing of human speech (**Cognitive Knowledge Level: Apply**)
CO6	Design, Develop, and Implement innovative ideas on speech processing concepts and techniques. **(Cognitive Knowledge Level: Create)**

Program Outcomes

Graduates of this program will be able to demonstrate the following attributes.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams.

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialisation of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards.

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of the-art tools to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects.

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1			✓			✓	
CO 2			✓		✓	✓	
CO 3			✓		✓	✓	
CO 4			✓			✓	
CO 5			✓			✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	50-80%
Analyse	20-40%
Evaluate	-
Create	-

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective college.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Explain the various vocal organs and how they produce speech phones. 2. Draw and explain the transducer for z-devoicing rule.

Course Outcome 2 (CO2):

1. Write brief notes on the noisy channel model of pronunciation and spelling.

2. What is meant by the minimum edit distance between two strings. Calculate the minimum edit distance between the words 'kitten' and 'sitting'.

Course Outcome 3 (CO3):

1. Illustrate the concept of pronunciation dictionaries with an example. 2. Discuss the various phonological aspects of prosody in speech.

Course Outcome 4 (CO4):

1. Give a brief overview of the Bayesian model of speech recognition. 2. List and briefly explain the various parameters needed to define an HMM.

Course Outcome 5 (CO5):

1. Highlight the steps involved in extraction of spectral features from sound wave. 2. List the different approaches for calculating probabilities of acoustic feature vectors.

Model Question paper

Total Pages: 2

Reg No.:_______________ Name:_________________________

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS021

Course Name: Speech Processing

Max. Marks: 60 Duration: 2.5 Hours

Branch : Computational Linguistics

PART A

Answer all questions in PART A. Each question carries 5 marks

Q. No		Marks
1	Write the lexical entry for the pronunciation of the English past tense suffix -d, and the two-level rules that express the difference in its pronunciation.	(5)
2	State the important three concerns while performing pronunciation dictionary lookup.	(5)
3	Define weighted automata. Give weighted automata for the word 'tomato'.	(5)
4	What are hidden Markov models? State the parameters needed to define HMM.	(5)

5		Summarize the process of feature extraction of sound waves.	(5)

PART B

Answer any five full questions in PART B. Each full question carries 7 marks

6		Design a transducer for i-insertion rule by writing appropriate phonological rules.	(7)
7		Write notes on the place and manner of articulation of consonants.	(7)
8	a)	Write notes on the Bayesian model for spelling correction.	(4)
	b)	State the forward algorithm. Also, explain how it is useful in computing likelihoods.	(3)
9		Using the minimum edit distance method, find whether the word 'drive' is closer to 'brief' or to 'divers'.	(7)
10		Explain the use of the Viterbi algorithm in speech recognition process.	(7)
11		Writes notes on speech recognition architecture.	(7)
12		Discuss how a speech recognition system can be trained.	(7)

****

Syllabus

Module	Syllabus Content	No.of Lecture Hours (38)
I	**Computational Phonology** - Articulatory Phonetics – Production and Classification of Speech Sounds – Vocal Organs - Consonants – Place of Articulation; Consonants – Manner of Articulation; Vowels - Phoneme and Phonological Rules; Phonological Rules and Transducers	8
II	**Speech Synthesis** - Mapping Text to Phonemes for TTS Pronunciation Dictionaries-Text Analysis-FST based pronunciation lexicon - Prosody in TTS – Phonological aspects of Prosody – Phonetic aspects of Prosody – Prosody in speech synthesis	8
III	**Models of Spelling and Pronunciation** - Spelling errors - Spelling Error Patterns-Detecting Nonword Errors - Probabilistic models of spelling - Bayesian method to spelling – Minimum Edit Distance - The Bayesian Method for Pronunciation-Decision Tree Models of Pronunciation Variation – Weighted Automata and Segmentation	10
IV	**Speech Recognition** - Speech Recognition Architecture – Bayesian Model of Speech Recognition - Hidden Markov Models - Viterbi Algorithm – Advanced Methods for Decoding - A* Decoding	6
V	**Acoustic processing of speech** - Sound Waves - Interpreting a Waveform – Spectra – Feature Extraction - Computing Acoustic Probabilities - Gaussian Models - Neural Net Models - Training a Recognizer	6

Course Plan

No	Topic	No. of Lectures
1	**Computational Phonology (8 Hours)**	
1.1	Articulatory Phonetics - Introduction	1
1.2	Production and Classification of Speech Sounds	1
1.3	Vocal Organs	1
1.4	Consonants – Place of Articulation	1
1.5	Consonants – Manner of Articulation	1
1.6	Articulation of vowels	1
1.7	Phoneme and Phonological Rules	1
1.8	Phonological Rules and Transducers	1
2	**Speech Synthesis (8 Hours)**	
2.1	Mapping Text to Phonemes for TTS	1
2.2	Pronunciation Dictionaries	1
2.3	Text Analysis	1
2.4	FST based Pronunciation Lexicon	1
2.5	Prosody in TTS	1
2.6	Phonological Aspects of Prosody	1

2.7	Phonetic Aspects of Prosody	1
2.8	Prosody in Speech Synthesis	1
3	**Models of Spelling and Pronunciation (10 Hours)**	
3.1	Spelling errors - Spelling Error Patterns	1
3.2	Detecting Nonword Errors	1
3.3	Probabilistic Models of Spelling	1
3.4	Bayesian Method to Spelling	1
3.5	Minimum Edit Distance	1
3.6	The Bayesian Method for Pronunciation	1
3.7	Decision Tree Models of Pronunciation Variation	1
3.8	Weighted Automata	1
3.9	Computing Likelihoods from Weighted Automata - The Forward Algorithm	1
3.10	Segmentation	1
4	**Speech Recognition (6 Hours)**	
4.1	Speech Recognition Architecture	1
4.2	Bayesian Model of Speech Recognition	1
4.3	Hidden Markov Models	1
4.4	Viterbi Algorithm	1
4.5	Advanced Methods for Decoding	1

4.6	A* Algorithm	1
5	**Acoustic processing of speech (6 Hours)**	
5.1	Sound Waves - Interpreting a Waveform	1
5.2	Spectra Analysis	1
5.3	Feature Extraction from Waveforms and Spectra	1
5.4	Computing Acoustic Probabilities - Gaussian Models	1
5.5	Computing Acoustic Probabilities - Neural Net Models	1
5.6	Training a Recognizer	1

Reference Books

1. Jurafsky, D. and J. H. Martin, Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Upper Saddle River, NJ: Prentice-Hall, 2000

2. Claudio Becchetti and Lucio Prina Ricotti, "Speech Recognition", John Wiley and Sons, 1999

3. Frederick Jelinek, "Statistical Methods of Speech Recognition", MIT Press, 1997

4. Ben Gold and Nelson Morgan, "Speech and Audio Signal Processing, Processing and Perception of Speech and Music", Wiley- India Edition, 2006

221ECS022	INFORMATION THEORY	CATEGORY	L	T	P	CREDIT
		PROGRAM ELECTIVE 2	3	0	0	3

Preamble: The course introduces the mathematical and fundamental notions of information theory that play a significant role in building modern communication systems. It covers entropy, mutual information, source coding, channel coding, continuous sources and channels and rate distortion theory. The course enables the learners to effectively choose appropriate source codes and channel codes according to different applications.

Course Outcomes:

After the completion of the course the student will be able to

CO 1	Compare different types of entropy and use the concept of mutual information. (**Cognitive Knowledge Level: Apply**)
CO 2	Design source codes and appreciate the use of Shannon's source coding theorem. (**Cognitive Knowledge Level: Apply**)
CO 3	Design channel codes and appreciate the use of Shannon's channel coding theorem. (**Cognitive Knowledge Level: Apply**)
CO 4	Demonstrate the notions of continuous sources and channels. (**Cognitive Knowledge Level: Apply**)
CO 5	Elaborate on the various aspects of the rate distortion theorem. (**Cognitive Knowledge Level: Apply**)
CO 6	Design, Develop, and Implement applications using Information theory concepts and techniques. **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams.

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards.

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓		✓		
CO 2	✓		✓	✓	✓		
CO 3	✓		✓	✓	✓		
CO 4	✓		✓		✓		
CO 5	✓						
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	50-80%
Analyse	20-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation: 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module

of which student should answer any five. Each question can carry 7 marks. Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. A source produces 4 symbols with probabilities 1/2, 1/2, 1/8, and 1/8. Find the information content of each symbol.

2. A zero memory source has a source alphabet, S = {s1, s2, s3} with P = {0.5, 0.3, 0.2}. Find the entropy of the source.

3. Given a binary source with two symbols x1 and x2. Given x2 is twice as long as x1 and half as probable. The duration of x1 is 0.3 seconds. Calculate the information rate of the source.

Course Outcome 2 (CO2)

1. Consider a [7,4] linear block code with parity check matrix

$$H = \begin{matrix} 1\,0\,1\,1\,1\,0\,0 \\ 1\,1\,0\,1\,0\,1\,0 \\ 0\,1\,1\,1\,0\,0\,1 \end{matrix}$$

a) Construct code words for the [7,4] code.

b) Show that this code is a Hamming code.

2. Consider a source with 8 alphabets, a to h with respective probabilities 0.2, 0.2, 0.18, 0.15, 0.12, 0.08, 0.05 and 0.02. Construct a minimum redundancy code and determine the code efficiency.

3. The parity matrix for a (6,3) systematic linear block code is given by

$$P = \begin{matrix} 1\,1\,0 \\ 1\,0\,1 \\ 0\,1\,1 \end{matrix}$$

(i) Find all code words. (ii) Find generator and parity check matrix.

Course Outcome 3(CO3):

1. State and prove Shannon's channel coding theorem.

2. Derive the capacity of binary symmetric channel.

3. State and prove source channel theorem.

Course Outcome 4 (CO4):

1. Derive the differential entropy of a normal distribution.

2. Derive the chain rule for differential entropy.3.

Course Outcome 5 (CO5):

1. State and prove the converse to the rate distortion theorem.

2. Evaluate the rate distortion function for a binary source.

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES : 2**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 221ECS022

Course Name: Information Theory

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Derive the relation between entropy and mutual information.

2. State Kraft-McMillan inequality.

3. Explain channel coding theorem.

4. Discuss about Gaussian channels.

5. Define rate distortion theorem. **(5x5 =25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6. Differentiate between joint entropy and conditional entropy. **(7)**

7. Suppose one has n coins, among which there may or may not be one counterfeit coin. If there is a counterfeit coin, it may be either heavier or lighter than the other coins. The coins are to be weighed by a balance. **(7)**

a) Find an upper bound on the number of coins n so that k weighings will find the counterfeit coin (if any) and correctly declare it to be heavier or lighter. b) What is the coin weighing strategy for k = 3 weighings and 12 coins?

8. Find a binary Huffman code for the source emitting symbols with probabilities 0.49, 0.14, 0.14, 0.07, 0.04, 0.02, 0.02, 0.01. Also find the code efficiency and redundancy. **(7)**

9. Explain with the help of a neat diagram, discrete memoryless channel with feedback. **(7)**

10. Consider the random variable. **(7)**

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}$$

a) Find a binary Huffman code for X.

b) Find the expected code length for this encoding.

c) Find a ternary Huffman code for X.

11. Consider the discrete memoryless channel Y = X + Z (mod 11), where **(7)**

$$Z = \begin{pmatrix} 1, & 2, & 3 \\ 1/3, & 1/3, & 1/3 \end{pmatrix}$$ and X∈{0,1,2, …, 10}. Assume that Z is independent of X.

a) Find the capacity.

b) What is the maximizing p*(x)?

12. Consider a source X uniformly distributed on the set { 1,2, . . . , m}. Find the rate distortion function for this source with Hamming distortion, i.e., **(7)**

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x}, \\ 1 & \text{if } x \neq \hat{x}. \end{cases}$$

Syllabus

Entropy, lossless source coding, Huffman code, Shannon's source coding theorem, Shannon's channel coding theorem, continuous sources and channels, rate distortion theory.

	Syllabus	
Module	**Content**	**Hours**
1	**Introduction to Entropy**: Entropy- Memoryless sources - Markov sources Entropy of a discrete random variable - joint, conditional and relative entropy - mutual Information and conditional mutual information - Chain relation for entropy, relative entropy and mutual information	12
2	**Lossless source coding** - Uniquely decodable codes - Instantaneous codes Kraft's inequality - Optimal codes - Huffman code- Shannon's Source Coding Theorem.	7
3	**Channel coding** - Shannon's Channel Coding Theorem and its converse - Channels with feedback - Joint source channel coding Theorem.	7
4	**Continuous Sources and Channels**: Continuous Sources and Channels - Differential Entropy - Joint, relative and conditional differential entropy – Mutual information- Waveform channels- Gaussian channels.	7
5	**Rate Distortion Theory**: Introduction - Rate Distortion Function - Properties - Continuous Sources and Rate Distortion measure - Rate Distortion Theorem – Converse – Information Transmission Theorem - Rate Distortion Optimization.	7

Course Plan

No	Topic	No. of Lectures (40 hours)
1	**Module 1 (Introduction to entropy) (12 hrs)**	
1.1	Entropy - Memoryless sources	2
1.2	Markov sources	2
1.3	Entropy of a discrete random variable	2
1.4	Joint, conditional and relative entropy	2
1.5	Mutual information and conditional mutual information	2
1.6	Chain relation for entropy, relative entropy and mutual information	2
2	**Module 2 (Lossless source coding) (7 hrs)**	
2.1	Uniquely decodable codes	1
2.2	Instantaneous codes	1
2.3	Kraft's inequality	1
2.4	Optimal codes - Huffman code	2
2.5	Shannon's Source Coding Theorem	2
3	**Module 3 (Channel coding) (7 hrs)**	
3.1	Introduction to channel coding	1
3.2	Shannon's Channel Coding Theorem and its converse	2
3.3	Channels with feedback	2

3.4	Joint source channel coding Theorem	2
4	**Module 4 (Continuous Sources and Channels) (7 hrs)**	
4.1	Continuous Sources and Channels	2
4.2	Differential Entropy	1
4.3	Joint, relative and conditional differential entropy	1
4.4	Mutual information	1
4.5	Waveform channels	1
4.6	Gaussian channels	1
5	**Module 5 (Rate Distortion Theory) (7 hrs)**	
5.1	Introduction to rate distortion theory	1
5.2	Rate Distortion Function – Properties	1
5.3	Continuous Sources and Rate Distortion measure	1
5.4	Rate Distortion Theorem	1
5.5	Converse of rate distortion theorem	1
5.6	Information Transmission Theorem	1
5.7	Rate Distortion Optimization.	1

References

1. T. Cover and Thomas, Elements of Information Theory, John Wiley & Sons.

2. Robert Gallager, Information Theory and Reliable Communication, John Wiley & Sons.

3. R. J. McEliece, The theory of information & coding, Addison Wesley Publishing Co.

4. T. Bergu, Rate Distortion Theory a Mathematical Basis for DataCompression PH Inc.

5. R. W. Hamming. Coding and Information Theory. Prentice Hall Inc.

221LCS001	ADVANCED MACHINE LEARNING LAB	CATEGORY	L	T	P	Credit
		Laboratory 1	0	0	2	2

Preamble: Study of the course enables the learners to make use of the machine learning concepts and algorithms to derive data insights. The course provides exposure to the design and implementation aspects of machine learning algorithms such as decision trees, regression, naive bayes algorithm, clustering algorithms and artificial neural network. This helps the students to develop machine learning based solutions to real world problems.

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

CO#	Course Outcomes
CO1	Apply modern machine learning notions in predictive data analysis (**Cognitive Knowledge Level: Apply**)
CO2	Analyze the range of machine learning algorithms along with their strengths and weaknesses (**Cognitive Knowledge Level: Analyze**)
CO3	Design and develop appropriate machine learning models to solve real world problems. (**Cognitive Knowledge Level: Analyze**)
CO4	Build predictive models from data and analyze their performance (**Cognitive Knowledge Level: Create**)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and developmentwork in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	✓	✓	✓	✓	✓	✓	
CO2	✓	✓	✓	✓	✓	✓	
CO3	✓	✓	✓	✓	✓	✓	
CO4	✓	✓	✓	✓	✓	✓	

Continuous Internal Evaluation Pattern:

The laboratory courses will be having only Continuous Internal Evaluation and carries 100 marks.

Final assessment shall be done by two examiners; one examiner will be a senior faculty from the same department.

Continuous Evaluation : 60 marks

Final internal assessment : 40 marks

Lab Report:

All the students attending the Lab should have a Fair Report. The report should contains details of experiment such as Objective, Algorithm/Design, Description, Implementation, Analysis, Results, and Outcome. The report should contain a print out of the respective code with inputs addressing all the aspects of the algorithm described and corresponding outputs. All the experiments noted in the fair report should be verified by the faculty regularly. The fair report, properly certified by the faculty, should be produced during the time of the final assessment.

Syllabus

Decision tree (ID3), Naïve bayesian classifier , Bayesian network, Expectation Maximization (EM) algorithm,K-means algorithm, K-nearest neighbor, Regression, Cross validation, Support Vector Machine (SVM), Artificial neural network, Backpropagation algorithm, Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM), Google colab.

Practice Questions

1. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
2. Write a program to implement the naïve bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
3. Assuming a set of documents that need to be classified, use the naïve bayesian Classifier model to perform this task. Calculate the accuracy, precision, and recall for your data set.
4. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Python ML library classes/API.
5. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Python ML library classes/API in the program.
6. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Python ML library classes can be used for this problem.
7. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.
8. Write a program to implement 5-fold cross validation on a given dataset. Compare the accuracy, precision, recall, and F-score for your data set for different folds.

9. Implement SVM/Softmax classifier for CIFAR-10 dataset: (i) using KNN, (ii) using 3 layer neural network.
10. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
11. Image Captioning with Vanilla RNNs .
12. Image Captioning with LSTMs.
13. Familiarisation of cloud based computing like Google colab.

References:

1. Jiawei Han, Micheline Kamber, Jian Pei. Data Mining Concepts and Techniques, Third Edition. Morgan Kaufmann.

2. Christopher M. Bishop. Pattern recognition and machine learning. Springer 2006.

3. Ethem Alpaydin, Introduction to Machine Learning, 2nd edition, MIT Press 2010.

4. Mohammed J. Zaki and Wagner Meira, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, First South Asia edition, 2016.

5. Goodfellow, I., Bengio,Y., and Courville, A., Deep Learning, MIT Press, 2016.

6. Neural Networks and Deep Learning, Aggarwal, Charu C., c Springer International Publishing AG, part of Springer Nature 2018

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
221RGE100	RESEARCH METHODOLOGY & IPR	General Course	2	0	0	2

Preamble:

This course introduces the strategies and methods related to scientific research. The students are also trained in the oral presentation with visual aids and writing technical thesis/reports/research papers. The salient aspects of publication and patenting along with the crucial role of ethics in research is discussed.

Course Outcomes

After the completion of the course the student will be able to

CO 1	Approach research projects with enthusiasm and creativity.
CO 2	Conduct literature survey and define research problem
CO 3	Adopt suitable methodologies for solution of the problem
CO 4	Deliver well-structured technical presentations and write technical reports.
CO 5	Publish/Patent research outcome.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✔	✔				✔	
CO 2	✔	✔				✔	
CO 3	✔	✔				✔	
CO 4	✔	✔				✔	
CO 5	✔	✔				✔	
CO 6	✔	✔				✔	

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70 %
Analyse	30 %
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Course based task: 15 marks

Some sample course based tasks that can be performed by the student given below.

- *Conduct a group discussion based on the good practices in research.*
- *Conduct literature survey on a suitable research topic and prepare a report based on this.*

Seminar: 15 marks

Test paper: 10 marks

End Semester Examination Pattern:

Total Marks: 60

The examination will be conducted by the respective college with the question provided by the University. The examination will be for 150 minutes and contain two parts; Part A and Part B. Part A will contain 6 short answer questions with 1 question each from modules 1 to 4, and 2 questions from module 5. Each question carries 5 marks. Part B will contain only 1 question based on a research article from the respective discipline and carries 30 marks. The students are to answer the questions based on that research article.

Model Question paper

QP Code:		**Total Pages:**
Reg No.: _______________		**Name:** ___________________________

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
FIRST SEMESTER M. TECH DEGREE EXAMINATION, Month & Year

Course Code: 221RGE100
Course Name: RESEARCH METHODOLOGY & IPR

Max. Marks: 60		**Duration: 2.5 Hours**

PART A

		Answer all questions. Each question carries 5 marks	**Marks**
1		Discuss the salient recommendations for great research recommended by Richard Hamming in his famous talk “You and Your Research”	**30**
2		What are the characteristics of a good research question? Discuss with an example.	
3		Explain the difference between continuum, meso-scale and micro scale approaches for numerical simulation.	
4		Discuss any four rules of scientific writing.	
5		What are the requirements for patentability?	
6		What are the differences between copyright and trademark protection?	
		Read the given research paper and write a report that addresses the following issues **(The paper given can be specific to the discipline concerned**)	
7		What is the main research problem addressed?	**3**
8		Identify the type of research	**3**
9		Discuss the short comings in literature review if any?	**6**
10		Discuss appropriateness of the methodology used for the study	**6**
11		Discuss the significance of the study and summarize the important results and contributions by the authors	**6**
12		Identify limitations of the article if any.	**6**

Syllabus and Course Plan

No	Topic	No. of Lectures
1	**Introduction**	
1.1	Meaning and significance of research, Skills, habits and attitudes for research, Types of research,	1
1.2	Characteristics of good research, Research process	1
1.3	Motivation for research: Motivational talks on research: "You and Your Research"- Richard Hamming	1
1.4	Thinking skills: Levels and styles of thinking, common-sense and scientific thinking, examples, logical thinking, division into sub-problems, verbalization and awareness of scale.	1
1.5	Creativity: Some definitions, illustrations from day to day life, intelligence versus creativity, creative process, requirements for creativity	1
2	**Literature survey and Problem definition**	
2.1	Information gathering – reading, searching and documentation, types of literature.	1
2.2	Integration of research literature and identification of research gaps	1
2.3	Attributes and sources of research problems, problem formulation, Research question, multiple approaches to a problem	1
2.4	Problem solving strategies – reformulation or rephrasing, techniques of representation, Importance of graphical representation, examples.	1
2.5	Analytical and analogical reasoning, examples, Creative problem solving using Triz, Prescriptions for developing creativity and problem solving.	1
3	**Experimental and modelling skills**	
3.1	Scientific method, role of hypothesis in experiment, units and dimensions, dependent and independent variables, control in experiment	1
3.2	precision and accuracy, need for precision, definition, detection, estimation and reduction of random errors, statistical treatment of data, definition, detection and elimination of systematic errors,	1
3.3	Design of experiments, experimental logic, documentation	1

3.4	Types of models, stages in modelling, curve fitting, the role of approximations, problem representation, logical reasoning, mathematical skills.	1
3.5	Continuum/meso/micro scale approaches for numerical simulation, Two case studies illustrating experimental and modelling skills.	1
4	**Effective communication - oral and written**	
4.1	Examples illustrating the importance of effective communication, stages and dimensions of a communication process.	1
4.2	Oral communication –verbal and non-verbal, casual, formal and informal communication, interactive communication, listening, form, content and delivery, various contexts for speaking- conference, seminar etc.	1
4.3	Guidelines for preparation of good presentation slides.	1
4.4	Written communication – Rules of scientific writing, form, content and language, layout, typography and illustrations, nomenclature, reference and citation styles, contexts for writing – paper, thesis, reports etc. Tools for document preparation-LaTeX.	1
4.5	Common errors in typing and documentation	1
5	**Publication and Patents**	
5.1	Relative importance of various forms of publication, Choice of journal and reviewing process, Stages in the realization of a paper.	1
5.2	Research metrics-Journal level, Article level and Author level, Plagiarism and research ethics	1
5.3	Introduction to IPR, Concepts of IPR, Types of IPR	1
5.4	Common rules of IPR practices, Types and Features of IPR Agreement, Trademark	1
5.5	Patents- Concept, Objectives and benefits, features, Patent process – steps and procedures	2

Reference Books

1. E. M. Phillips and D. S. Pugh, "How to get a PhD - a handbook for PhD students and their supervisors", Viva books Pvt Ltd.
2. G. L. Squires, "Practical physics", Cambridge University Press
3. Antony Wilson, Jane Gregory, Steve Miller, Shirley Earl, Handbook of Science Communication, Overseas Press India Pvt Ltd, New Delhi, 1st edition 2005
4. C. R. Kothari, Research Methodology, New Age International, 2004
5. Panneerselvam, Research Methodology, Prentice Hall of India, New Delhi, 2012.

6. Leedy P. D., Practical Research: Planning and Design, McMillan Publishing Co.
7. Day R. A., How to Write and Publish a Scientific Paper, Cambridge University Press, 1989.
8. William Strunk Jr., Elements of Style, Fingerprint Publishing, 2020
9. Peter Medawar, 'Advice to Young Scientist', Alfred P. Sloan Foundation Series, 1979.
10. E. O. Wilson, Letters to a Young Scientist, Liveright, 2014.
11. R. Hamming, You and Your Research, 1986 Talk at Bell Labs.

SEMESTER II

Discipline:COMPUTER SCIENCE AND ENGINEERING

Stream : CS2

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222TCS100	ADVANCED DATA STRUCTURES AND ALGORITHMS	DISCIPLINE CORE 2	3	0	0	3

Preamble: The course introduces advanced data structures and algorithms in different domains. The goal of this course is to provide a solid background in the design and analysis of the major classes of algorithms. The course helps the learners to develop their own versions for a given computational task and to compare and contrast their performance.

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

CO 1	Analyse the relevance of amortized analysis and applications. (Cognitive Level: Apply)
CO 2	Illustrate string matching algorithms. (Cognitive Level: Apply)
CO 3	Illustrate advanced data structures like Binomial heap, Fibonacci heap, Disjoint set and string matching algorithms. (Cognitive Level: Apply)
CO 4	Illustrate network flow algorithms and applications. (Cognitive Level: Apply)
CO 5	Make use of probabilistic algorithms and approximation algorithms in computing. (Cognitive Level: Apply)
CO 6	Design, develop and implement software using advanced data structures and algorithms. (Cognitive Level: Create)

* The COs shown are only indicative. For each course, there can be 4 to 6 COs.

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓		✓	
CO 2	✓		✓	✓		✓	
CO 3	✓		✓	✓		✓	
CO 4	✓		✓	✓		✓	
CO 5	✓		✓	✓		✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	80%
Analyse	20%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation : 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/Quiz : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted.

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Explain how the accounting method of amortized analysis can be applied to stack operations.

2. Suppose we perform a sequence of n operations on a data structure in which the i^{th} operation costs i if i is an exact power of 2, and 1 otherwise. Use aggregate analysis to determine the amortized cost per operation.

3. What is the total cost of executing n of the stack operations PUSH, POP, and MULTIPOP, assuming that the stack begins with s_0 objects and finishes with s_n objects? Use potential method.

Course Outcome 2 (CO2)

1. Use an aggregate analysis to show that the running time of KMP-MATCHER is $\theta(n)$.

2. Working modulo q = 11, how many spurious hits does the Rabin-Karp matcher encounter in the text T = 3141592653589793 when looking for the pattern P = 26?

3. Compute the prefix function π for the pattern ababbabbabbababbabb

Course Outcome 3(CO3):

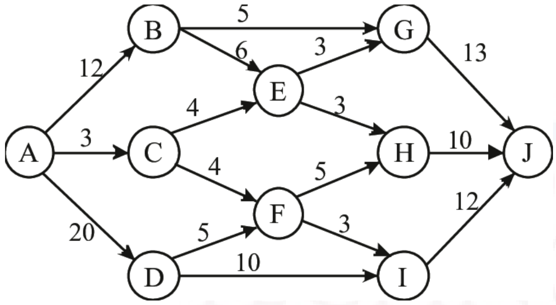
1. Analyse the time complexity of decrease-key operation of Fibonacci heap.

2. Illustrate extract-min operation of Binomial heap.

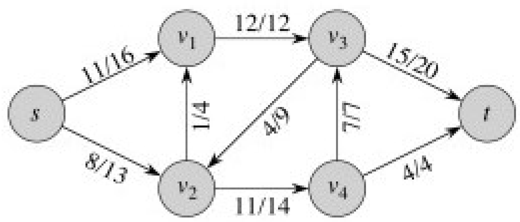
3. Explain the heuristics used in disjoint set data structure to improve the running time.

Course Outcome 4 (CO4):

1. Show the execution of the Edmonds-Karp algorithm on the given flow network (source: A and sink: J).



2. In the following figure, compute flow across the cut ({s, v_2, v_4}, {v_1, v_3, t}). What is the capacity of this cut?



3. State and prove max flow min cut theorem.

Course Outcome 5 (CO5):

1. Illustrate Miller-Rabin primality testing method.

2. Explain probabilistic selection algorithm.

Course Outcome 6 (CO6):

1. Explain the approximation algorithm for subset sum problem.

2. Consider each of the following words as a set of letters: *{arid, dash, drain, heard, lost, nose, shun, slate, snare, thread}*. Show which set cover GREEDY-SET-COVER produces when we break ties in favor of the word that appears first in the dictionary.

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 222TCS100

Course Name: ADVANCED DATA STRUCTURES AND ALGORITHMS

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Explain accounting method of amortized analysis with a suitable example.

2. Explain the algorithm for uniting two binomial heaps and analyse the running time.

3. Maximum matching in a bipartite graph G corresponds to a maximum flow in its corresponding flow network G'. Comment on this statement. Explain how maximum flow problem can be used to solve maximum bipartite matching problem.

4. Explain thbe probabilistic algorithm for verifying matrix multiplication problem.

5. Explain the approximation algorithm for traveling salesperson problem.

(5x5=25)

Part B

(Answer any five questions. Each question carries 7 marks)

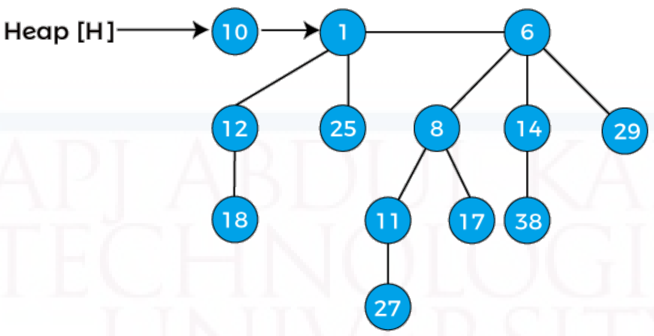
6. Describe Knuth-Morris-Pratt algorithm and illustrate using given text T = AABAACAADAABAABA and pattern P = AABA. **(7)**

7. (a) Using potential method, compute the amortized cost of incrementing a binary counter. **(3)**

(b) Suppose we perform a sequence of n operations on a data structure in which the i^{th} operation costs i if i is an exact power of 2, and 1 otherwise. Use accounting method of amortized analysis to determine the amortized cost per operation. **(4)**

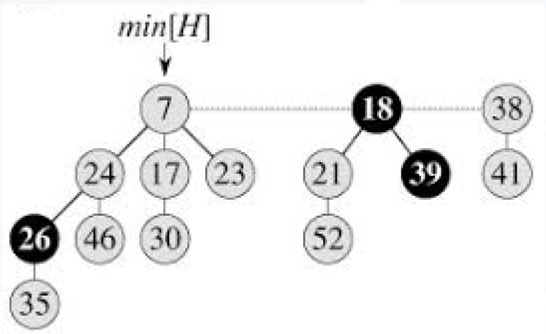
8. (a) Explain how disjoint set data structure is used to find connected components on an undirected graph. **(3)**

(b) Show the binomial heap that results when a node with key 11 is deleted from the binomial heap shown in figure. **(4)**

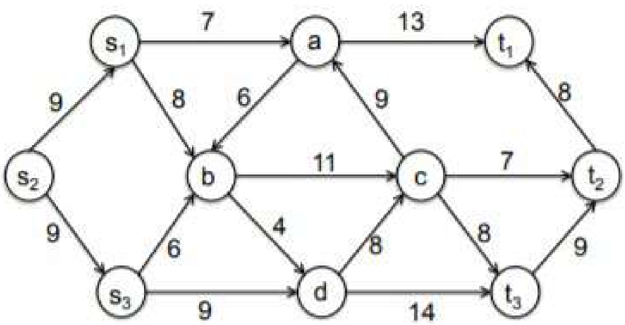


9. (a) Explain the structure of Fibonacci heap. **(2)**

(b) Apply extract minimum operation on the Fibonacci heap shown in figure and show the result. **(5)**



10 Describe Ford-Fulkerson algorithm and apply on the following network. Also obtain minimum cut across the network. **(7)**



11 (a) Apply Miller-Rabin algorithm to test whether the number 341 is prime or not. **(4)**

(b) Explain probabilistic quick sort algorithm. **(3)**

12 (a) Describe polynomial-time approximation scheme and fully polynomial-time approximation scheme. **(3)**

(b) Give an example of a graph for which APPROX-VERTEX-COVER always yields a suboptimal solution. **(4)**

Syllabus

Module – 1 (Amortized analysis and String matching)
Overview of asymptotic notations and complexity analysis, Amortized analysis – aggregate analysis, accounting method, potential method.
String matching – introduction, Rabin-Karp algorithm, Knuth-Morris-Pratt algorithm.

Module – 2 (Advanced data structures)
Overview of binary heap operations, Binomial tree and heap, Binomial heap operations, Fibonacci heap structure, Fibonacci heap operations, Disjoint set – overview, linked list representation, disjoint set forests.

Module – 3 (Network flow)
Network flow properties, examples, residual network, augmenting path, cut of network, maxflow-mincut theorem, Ford-Fulkerson algorithm, Edmonds-Karp algorithm, maximum bipartite matching.

Module – 4 (Probabilistic algorithms)
Introduction, types of probabilistic algorithms, Numerical algorithms – Numerical integration, Probabilistic counting, Monte-Carlo algorithms – Verifying matrix multiplication.
Number theory fundamentals – modular arithmetic, modular exponentiation, Euler's Theorem and Fermat's Theorem, Primality testing – Miller-Rabin test.
Las Vegas algorithms – Probabilistic selection and quick sort.

Module – 5 (Approximation algorithms)
Introduction, Vertex-cover problem, Traveling-salesman problem, Set-covering problem, Subset-sum problem.

Course Plan

No	Topic	No. of Lectures (37)
1	**Module – 1 (Amortized analysis and String matching)**	
1.1	Overview of asymptotic notations and complexity analysis	1
1.2	Amortized analysis – aggregate analysis	1
1.3	accounting method	1
1.4	potential method	1
1.5	String matching – introduction	1
1.6	Rabin-Karp algorithm	1
1.7	Knuth-Morris-Pratt algorithm (1)	1
1.8	Knuth-Morris-Pratt algorithm (2)	1
2	**Module – 2 (Advanced data structures)**	
2.1	Overview of binary heap operations	1
2.2	Binomial tree and heap	1

2.3	Binomial heap operations (1)	1
2.4	Binomial heap operations (2)	1
2.5	Fibonacci heap structure	1
2.6	Fibonacci heap operations (1)	1
2.7	Fibonacci heap operations (2)	1
2.8	Disjoint set – overview, linked list representation	1
2.9	disjoint set forests	1
3	**Module – 3 (Network flow)**	
3.1	Network flow properties, examples	1
3.2	residual network, augmenting path, cut of network	1
3.3	maxflow-mincut theorem	1
3.4	Ford-Fulkerson algorithm	1
3.5	Edmonds-Karp algorithm	1
3.6	Maximum bipartite matching	1
4	**Module – 4 (Probabilistic algorithms)**	
4.1	Introduction, types of probabilistic algorithms	1
4.2	Numerical algorithms – Numerical integration, Probabilistic counting	1
4.3	Monte-Carlo algorithms – Verifying matrix multiplication	1
4.4	Number theory fundamentals – modular arithmetic, modular exponentiation	1
4.5	Euler's Theorem and Fermat's Theorem	1
4.6	Primality testing – Miller-Rabin test (1)	1
4.7	Primality testing – Miller-Rabin test (2)	1
4.8	Las Vegas algorithms – Probabilistic selection and quick sort	1
5	**Module – 5 (Approximation algorithms)**	
5.1	Introduction	1
5.2	Vertex-cover problem	1
5.3	Traveling-salesman problem	1
5.4	Set-covering problem	1
5.5	Subset-sum problem (1)	1
5.6	Subset-sum problem (2)	1

Reference Books

1. T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, “Introduction to Algorithms”, MIT Press, 3rd edition, 2009.

2. Gilles Brassard and Paul Bratley, “Fundamentals of algorithms”, Prentice-hall of India Private Limited, 2001.

3. Rajeev Motwani, Prabhakar Raghavan, “Randomized Algorithms”, Cambridge University Press, 2000.

4. Dexter C. Kozen, “The Design and Analysis of Algorithms”, Springer.

5. Jon Kleinberg and Eva Tardos, “Algorithm Design”, Pearson Education, 2006.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222TCS002	DEEP LEARNING	PROGRAM CORE 3	3	0	0	3

Preamble: This course introduces the core concepts of deep learning and also provides an insight into recent developments in the field. The concepts covered in the course include Neural Network Optimization techniques, Regularization, Convolutional Neural networks, Recurrent Neural Networks, Word Embedding and Transformers. This course helps the students to develop solutions to real world applications using deep learning techniques.

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

CO 1	Analyse the challenges in learning of Neural Networks and develop solutions to overcome the issues **(Cognitive knowledge level: Apply)**
CO 2	Construct convolutional neural networks for deep learning applications **(Cognitive knowledge level: Apply)**
CO 3	Make use of recurrent neural network and its variants in relevant application areas **(Cognitive knowledge level: Analyze)**
CO 4	Apply the deep learning techniques in natural language based applications **(Cognitive knowledge level: Analyze)**
CO 5	Distinguish the transformer architecture from earlier architectures **(Cognitive knowledge level: Analyze)**
CO6	Design, develop and implement solutions based on Deep Learning concepts and techniques **(Cognitive knowledge level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓	✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	
CO 5	✓		✓	✓	✓	✓	
CO6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	40
Analyse	20
Evaluate	Can be evaluated using miniprojects/assignments
Create	Can be evaluated using miniprojects/assignments

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation : 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/Quiz : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted.

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Apply the Adam optimization algorithm to an example of your choice and implement it using appropriate tools..

2. Explore the validity of the statement: “Computational complexity due to Batch normalization is proportional to the size of the mini batch”.

Course Outcome 2 (CO2)

1. Design and sketch a CNN for any application of your choice.

2. Differentiate between a CNN and an autoencoder in terms of architecture and applications.

3. Demonstrate the use of transfer learning in natural language processing.

Course Outcome 3(CO3):

1. Illustrate the workflow inside an RNN by using an example.

2. In comparison to an RNN, how many orders of excess computation are there in an LSTM?

3. Implement any sequence to sequence based application using RNN and LSTM. Analyse which one of these two is more relevant for your application.

Course Outcome 4 (CO4):

1. Provide schematic diagrams showing input, output and hidden layers along with their connectivity for the word2vec embedding method.

2. Analyze how good BERT is in the application area of automated question answering.

3. Derive the loss functions used for training of a GAN.

Course Outcome 5 (CO5):

1. Differentiate between self-attention and multi-head attention.

2. Justify the statement: “Vision transformer is an adaptation of the transformer architecture for computer vision applications' '.

3. Analyze the computational complexities involved in large scale pretraining of transformers.

Course Outcome 6 (CO6):

1. Develop a deep learning model as a solution to a real world problem and analyse its performance.

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 222TCS002

Course Name: DEEP LEARNING

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1.		Consider a simple perception f such that $f:R^4 \rightarrow R$ which uses sigmoid as its activation function. The inputs are X=[0.1 ,0.2, 0.3, 0.4] and the corresponding weights W=[−0.2, 0.1, 0.2, 0.3]. Compute the derivative of sigmoid of z(σ'(z)), where z is a linear combination of weights vector W and input vector X. Ignore the bias term.	
2.		Illustrate convolution and pooling operation with an example.	
3.		Explain the relevance of LSTM in the context of recurrent neural networks. .	
4.		Describe the word2vec technique of word embedding.	
5.		Illustrate the concept of self-attention in transformers.	**(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	a	Explain the concepts behind i) Early stopping ii) dropout iii) weight decay	**(4)**
	b	Using Adagrad-based gradient descent, find the new value of parameter θ_{t+1}, given that the old value θ_t=0.7, aggregated gradient $\Delta\theta_t$=0.9, gradient accumulation r_{t-1}=0.6, learning rate α=0.1 and small constant $\delta=10^{-8}$.	**(3)**
7.	a	The input to a CNN architecture is a color image of size 112x112x3. The first convolution layer comprises 64 kernels of size 5x5 applied with a stride of 2 and padding 0. What will be the number of parameters?	**(3)**
	b.	Explain how the Vanishing and Exploding gradient problem is addressed in ResNet.	(4)
8.		Explain the working of RNN and discuss how backpropagation through time is used in recurrent networks.	**(7)**
9.		Illustrate the GloVe technique of word embedding.	**(7)**
10.		Demonstrate the workflow of the BERT technique of word embedding.	**(7)**
11.		Justify the need for self-attention in the transformer architecture.	**(7)**
12.		Illustrate the architecture of a Vision transformer.	**(7)**

Syllabus

Module 1: Introduction to Deep Learning **(8 hours)**
Deep Learning vs traditional machine learning, Gradient Descent, Adam Optimization, Weight initialization strategies, Batch Normalization, Regularization techniques, Cross entropy loss function
Module 2: Convolutional Neural Networks **(9 hours)**
Convolution operation, CNN layers, Building a CNN model, Training a CNN, Deep Autoencoders,
Module 3: Recurrent Neural Networks **(8 hours)**
Basic architecture and variants, Backpropagation through time, LSTM, Deep recurrent neural networks, Machine Translation, Encoder Decoder architecture, Sequence to sequence learning, Beam search
Module 4: Text processing using deep learning **(8 hours)**
Word embeddings, word2vec, GloVe, Subword embedding, BERT

Module 5: Transformers (8 hours)
Bahdanau attention, Multi-head attention, Self-attention and positional encoding, Transformer architecture, Transformers for vision, Large-scale pretraining with transformers

Course Plan

No	Topic	No. of Lectures
1	**Introduction to Deep Learning**	
1.1	Deep Learning vs traditional machine learning	1
1.2	Gradient Descent	1
1.3	Challenges in Neural Network Optimization-Ill conditioning, local Minim, Saddle Points, Exploding Gradients	1
1.4	Adagrad,RMSProp, Adam Optimization	1
1.5	Weight initialization strategies	1
1.6	Batch Normalization	1
1.7	Regularization - Parameter Norm Penalties L2, L1 Regularization	1
1.8	Regularization -Dataset Augmentation, Noise Robustness, Early Stopping, Dropouts	1
2	**Convolutional Neural Networks**	
2.1	Convolution operation	1
2.2	Convolution -Sparse Interaction, Parameter sharing, Equi-varient Representation	1
2.3	Pooling, Convolutional Neural Networks, Parameters	1
2.4	Variants of basic convolution function- dilated convolution, Transpose Convolution, 1x1 convolution, 3D Convolution	1
2.5	Backpropagation in Convolutional Layers	1
2.6	Cross Entropy Loss Function, AlexNet	1
2.7	GoogleNet Architecture, Inception Module,	1
2.8	Autoencoders- sparse autoencoders	1
2.9	Denoising autoencoders	1
3	**Recurrent Neural Networks**	
3.1	Unfolding Graphs	1
3.2	Recurrent Neural Networks, Computing Gradients	1
3.3	Modeling Sequences Conditioned on Context with RNNs	1

3.4	Encoder-Decoder Sequence-to-Sequence Architecture	1
3.5	Deep recurrent neural networks, Recursive Neural Networks	1
3.6	Exploding and Vanishing Gradients	1
3.7	LSTM	
3.8	Beam search	1
4	**Word embedding, BERT**	
4.1	Word embeddings	1
4.2	word2vec-Skip-gram and Continuous Bag of words	1
4.3	Word Embedding with Global Vectors (GloVe)	1
4.4	Subword embedding	1
4.5	Hierarchical Softmax	1
4.6	Bidirectional Encoder Representations from Transformers (BERT)	1
4.7	Natural Language Processing: with pretraining using word embedding	1
4.8	Case studies - Sentiment Analysis	1
5	**Transformers**	
5.1	Bahdanau attention	1
5.2	Multi-head attention	1
5.3	Self-attention and positional encoding	1
5.4	Transformer architecture	1
5.5	Transformers for vision (Lecture 1)	1
5.6	Transformers for vision (Lecture 2)	1
5.7	Large-scale pretraining with transformers (Lecture 1)	1
5.8	Large-scale pretraining with transformers (Lecture 2)	1

Reference Books

1. Ian Goodfellow, YoshuaBengio and Aaron Courville, Deep Learning, Second edition, MITPress, 2016.

2. M. Gopal, Deep Learning, Pearson, 2022

3. Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola, Dive into Deep Learning, available online at d2l.ai,

COURSE CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222PCS100	MINI PROJECT	PROJECT	0	0	4	2

Mini project can help to strengthen the understanding of student's fundamentals through application of theoretical concepts and to boost their skills and widen the horizon of their thinking. The ultimate aim of an engineering student is to resolve a problem by applying theoretical knowledge. Doing more projects increases problem solving skills.

The introduction of mini projects ensures preparedness of students to undertake dissertation. Students should identify a topic of interest in consultation with PG Programme Coordinator that should lead to their dissertation/research project. Demonstrate the novelty of the project through the results and outputs.The progress of the mini project is evaluated based on three reviews, two interim reviews and a final review. A report is required at the end of the semester.

Evaluation Committee - Programme Coordinator, One Senior Professor and Guide.

Sl. No	Type of evaluations	Mark	Evaluation criteria
1	Interim evaluation 1	20	
2	Interim evaluation 2	20	
3	Final evaluation by a Committee	35	Will be evaluating the level of completion and demonstration of functionality/ specifications, clarity of presentation, oral examination, work knowledge and involvement
4	Report	15	the committee will be evaluating for the technical content, adequacy of references, templates followed and permitted plagiarism level(not more than 25%)
5	Supervisor/Guide	10	
	Total Marks	100	

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222LCS001	DEEP LEARNING LAB	LABORATORY 2	0	0	2	1

Preamble: This course provides a practical introduction to deep learning algorithms in Python. This course includes programming exercises in computer vision, time series, natural language processing and generative modelling. Upon completing the course, the student will acquire the skills necessary to develop applications using deep learning frameworks.

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

CO#	Course Outcomes
CO1	Implement deep learning techniques to solve problems in computer vision **(Cognitive Knowledge Level: Apply)**
CO2	Implement deep learning techniques to solve problems involving time series data **(Cognitive Knowledge Level: Apply)**
CO3	Implement deep learning techniques to solve problems in text processing **(Cognitive Knowledge Level: Apply)**
CO4	Implement deep learning techniques to develop generative modeling **(Cognitive Knowledge Level: Apply)**

Programme Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1		✓			✓	✓	✓
CO2		✓			✓	✓	✓
CO3		✓			✓	✓	✓
CO4		✓			✓	✓	✓

Continuous Internal Evaluation Pattern:

The laboratory courses will be having only Continuous Internal Evaluation and carries 100 marks. Final assessment shall be done by two examiners; one examiner will be a senior faculty from the same department.

Continuous Evaluation : 60 marks

Final internal assessment : 40 marks

Lab Report:

All the students attending the Lab should have a Fair Report. The report should contains details of experiment such as Objective, Algorithm/Design, Description, Implementation, Analysis, Results, and Outcome. The report should contain a print out of the respective code with inputs addressing all the aspects of the algorithm described and corresponding outputs. All the experiments noted in the fair report should be verified by the faculty regularly. The fair report, properly certified by the faculty, should be produced during the time of the final assessment.

Syllabus

1) Basic deep learning for computer vision
2) Advanced deep learning for computer vision
3) Deep learning for time series
4) Deep learning for text
5) Generative deep learning
6) Deep learning using transformers

Practice Questions

Syllabus

1. Basic image processing operations : Histogram equalization, thresholding, edge detection, data augmentation, morphological operations
2. Implement SVM/Softmax classifier for CIFAR-10 dataset: (i) using KNN, (ii) using 3 layer neural network
3. Study the effect of batch normalization and dropout in neural network classifier
4. Familiarization of image labelling tools for object detection, segmentation
5. Image segmentation using Mask RCNN, UNet, SegNet
6. Object detection with single-stage and two-stage detectors (Yolo, SSD, FRCNN, etc.)
7. Image Captioning with Vanilla RNNs, Image Captioning with LSTMs.
8. Chatbot using bi-directional LSTMs
9. Implement time series forecasting using suitable datasets.
10. Implement sequence to sequence learning.

Reference Books:

1. Deep Learning with Python, by François Chollet, Manning, 2021

SEMESTER II

PROGRAM ELECTIVE III

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS003	COMPUTER VISION	PROGRAMME ELECTIVE 3	3	0	0	3

Preamble: This course provides the basic concepts and advanced techniques in computer vision. The areas comprising the syllabus include modern CNN architectures, modern object detection architectures, and most recent architectures using transformers. On completion of this course, the student would have an insight into the latest deep learning techniques used in computer vision.

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

CO 1	Express the computer vision pipleline**(Cognitive knowledge level: Apply)**
CO 2	Explore Modern convolutional neural network architectures **(Cognitive knowledge level: Apply)**
CO 3	Apply Modern object detection architectures **(Cognitive knowledge level: Apply)**
CO 4	Explore Generative Adversarial Networks for computer vision **(Cognitive knowledge level: Apply)**
CO 5	Explore recent architectures for computer vision **(Cognitive knowledge level: Apply)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1			✓			✓	
CO 2			✓			✓	
CO 3	✓	✓	✓	✓	✓	✓	✓
CO 4			✓			✓	
CO 5			✓	✓			

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	30
Analyse	30
Evaluate	Can be evaluated through assignments and miniproject
Create	Can be evaluated through assignments and miniproject

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks
ii. Course based task / Seminar/ Data collection and interpretation : 15 marks
iii. Test paper (1 number) : 10 marks
Test paper shall include minimum 80% of the syllabus.
Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. List the elements of the computer vision pipeline.

2. Give a case study of any two applications of computer vision.

3. Illustrate the relevance of each stage of the computer vision pipeline.

Course Outcome 2 (CO2)

1. Illustrate the architecture of a CNN.

2. Explore the VGGNet architecture.

3. Compare the ResNet architecture with GoogleNet architecture.

Course Outcome 3(CO3):

1. Outline the general object detection framework.
2. Compare various evaluation metrics for object detection.
3. Show the workflow of the Yolo algorithm.

Course Outcome 4 (CO4):

1. Illustrate the working of a DCGAN.
2. Show the architecture of a Pix2PixGAN.
3. Outline the functional elements of an SRGAN.

Course Outcome 5 (CO5):

1. Demonstrate the main components of the Vision Transformer.
2. Show how transformer and GAN are combined in the TransGAN.
3. State the need for linear projection in a Vision Transformer.

Model Question Paper		
QP CODE:		
Reg No: _______________		
Name: _________________ PAGES : 4		
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY		
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR		
Course Code: 222ECS003		
Course Name: Computer Vision		
Max. Marks : 60 Duration: 2.5 Hours		
PART A		
Answer All Questions. Each Question Carries 5 Marks		
1.	Classify the various elements of computer vision.	
2.	Examine the peculiarities of the MobileNet architecture.	
3.	Justify the elements in a general object detection framework.	
4.	Illustrate any one application of GAN in computer vision.	

Model Question Paper		
5.	Sketch the innovations in the Yolov7 technique in comparison with older versions.	**(5x5=25)**
Part B **(Answer any five questions. Each question carries 7 marks)**		
6.	You are given a CNN having a total of 10 layers. Devise a way to apply transfer learning to this network.	**(7)**
7.	Illustrate the inner architecture of the residual block in ResNet architecture.	**(7)**

8.	Compare and contrast the GoogleNet and ResNet architectures.	**(7)**
9.	Using a diagram, demonstrate the data flow in the Single Shot Detector architecture.	**(7)**
10.	Express in detail the workflow in a Yolo architecture when a 13x13 grid is applied on the input image.	**(7)**
11.	State the equation for total loss in a neural style transfer architecture and justify each loss term with respect to its function in the architecture.	**(7)**
12.	In the Vision Transformer architecture, justify the need for a trainable linear projection and show the computational steps prior to this projection.	**(7)**

Syllabus

Module 1 Introduction to Computer Vision (7 hours)
Introduction to computer vision, Computer vision, Applications of computer vision, Computer vision pipeline, Classification, Transfer learning and various approaches, Open source dataset
Module 2 Advanced CNN architectures (9 hours)
CNN architecture and components, Image classification using CNNs, AlexNet, VGGNet, Inception andGoogLeNet, ResNet, MobileNet
Module 3 Object detection (10 hours)
General object detection framework, Object-detector evaluation metrics, Region-based convolutional neural networks, Faster R-CNN, Single-shot detector (SSD), YOLO, YOLOv3
Module 4 GAN applications (5 hours)
DCGAN Image-to-image translation (Pix2PixGAN) , Image super-resolution GAN (SRGAN), Neural style transfer, Visual embeddings
Module 5 Advanced architectures (9 hours) YoloV7 ,Vision Transformer ,TransGAN, GPV-1, Mobile-former

Course Plan (For 3 credit courses, the content can be for 40 hrs and for 2 credit courses, the content can be for 26 hrs. The audit course in third semester can have content for 30 hours).

No	Topic	Hours
1	**Introduction to computer vision**	
1.1	Computer vision	1
1.2	Applications of computer vision	1
1.3	Applications of computer vision	1
1.4	Computer vision pipeline	1
1.5	Classification	1
1.6	Transfer learning and various approaches	1
1.7	Open source datasets	1
2	**Advanced CNN architectures**	
2.1	CNN architecture and components,	1
2.2	Image classification using CNNs	1
2.3	Image classification using CNNs	1
2.4	AlexNet	1
2.5	VGGNet	1
2.6	Inception and GoogLeNet	1
2.7	Inception and GoogLeNet	1

2.7	ResNet	1
2.8	MobileNet	1
3	**Object detection**	
3.1	General object detection framework	1
3.2	Object-detector evaluation metrics	1
3.3	Region-based convolutional neural networks	1
3.4	Faster R-CNN	1
3.5	Single-shot detector (SSD)	1
3.6	Single-shot detector (SSD)	1
3.7	YOLO	1
3.8	YOLO	1
3.9	YOLOv3	1
3.10	YOLOv3	1
4	**GAN applications in computer vision**	
4.1	DCGAN	1
4.2	Image-to-image translation (Pix2PixGAN)	1
4.3	Image super-resolution GAN (SRGAN)	1
4.4	Neural style transfer	1
4.5	Visual embeddings	1

5	**Advanced architectures**	
5.1	YoloV7	1
5.2	YoloV7	1
5.3	Vision Transformer	1
5.4	Vision Transformer	1
5.5	TransGAN	1
5.6	TransGAN	1
5.7	GPV-1	1
5.8	Mobile-former	1
5.9	Mobile-former	1

Reference Books

1. Deep Learning for Vision Systems, Mohamed Elgendy, Manning, 2020 **Reference Papers**

2. Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," https://doi.org/10.48550/arXiv.2207.02696, 2022.

3. Alexey Dosovitskiy et al, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", https://doi.org/10.48550/arXiv.2010.11929, 2020.

4. Yifan Jiang, Shiyu Chang, Zhangyang Wang, "TransGAN: Two Pure Transformers Can Make One Strong GAN, and That Can Scale Up", https://doi.org/10.48550/arXiv.2102.07074, 2021.

5. Tanmay Gupta, Amita Kamath, Aniruddha Kembhavi, Derek Hoiem, "Towards General Purpose Vision Systems", https://doi.org/10.48550/arXiv.2104.00743, 2021.

6. Yinpeng Chen, Xiyang Dai, Dongdong Chen, Mengchen Liu, Xiaoyi Dong, Lu Yuan, Zicheng Liu; "Mobile-Former: Bridging MobileNet and Transformer," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 5270-5279. https://openaccess.thecvf.com/content/CVPR2022/papers/Chen_Mobile Former_Bridging_MobileNet_and_Transformer_CVPR_2022_paper.pdf

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS012	BLOCKCHAIN TECHNOLOGIES	PROGRAMME ELECTIVE 3	3	0	0	3

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 and implement the cryptographic building blocks of blockchain technology. **(Cognitive Knowledge Level: Apply)**
CO2	Make use of the concepts of blockchain technology. **(Cognitive Knowledge Level: Apply)**
CO3	Summarize the classification of consensus algorithms. (**Cognitive Knowledge Level: Understand)**
CO4	Illustrate the concepts of first decentralized cryptocurrency bitcoin. (**Cognitive Knowledge Level: Apply)**
CO5	Implement smart contracts and its use cases. **(Cognitive Knowledge Level: Apply)**
CO6	Develop simple applications using Solidity language on Ethereum platform. **(Cognitive Knowledge Level: Apply)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and developmentwork in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓		✓	
CO 2	✓		✓	✓		✓	
CO 3	✓		✓	✓		✓	
CO 4	✓		✓	✓		✓	
CO 5	✓		✓	✓		✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	80%
Analyze	20%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii.Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.
Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:
The end semester examination will be conducted by the respective College.
There will be two parts; Part A and Part B.
Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (shall be useful in the testing of overall achievement and maturity of thestudents in a course, through long answer questions relating totheoretical/practical knowledge, derivations, problem solving and quantitativeevaluation), with minimum one question from each module of which studentshould answer any five. Each question can carry 7 marks
Total duration of the examination will be 150 minutes.
Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

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. Illustrate 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. Implement 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 : 4**
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY			
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR			
Course Code BLOCKCHAIN TECHNOLOGIES			
Course Name: 222ECS012			
Max. Marks : 60			**Duration: 2.5 Hours**
PART A			
Answer All Questions. Each Question Carries 5 Marks			
1.		Explain how hash functions are used to build Merkle trees in blockchain.	
2.		Explain the benefits, features and limitations of blockchain.	
3.		What is the role of a Bitcoin miner? Explain the mining algorithm used in Bitcoin with the help of a flowchart.	
4.		Explain the design process of decentralized applications with diagrams.	
5.		Define block difficulty. Explain how block difficulty is adjusted in Ethereum blockchain network.	**(5x5=25)**
Part B			
(Answer any five questions. Each question carries 7 marks)			
6.	(a)	Explain the design of SHA-256 and its compression function using a diagram.	**(7)**
7.	(a)	Explain the concept of Gas in Ethereum. Explain how transaction cost can be calculated in an Ethereum blockchain network.	**(7)**

Model Question Paper			
8.	(a)	Explain consensus mechanisms used in blockchain. List out any six consensus algorithms used in the context of blockchain.	**(7)**
9.	(a)	Show how Practical Byzantine Fault Tolerance can achieve consensus in the presence of Byzantine faults.	**(7)**
10.	(a)	Illustrate how blockchain technology can be implemented in finance sector.	**(7)**
11.	(a)	Using Solidity language, create a simple bank contract that allows a user to deposit, withdraw and view balance.	**(7)**

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.

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

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS013	IMAGE AND VIDEO ANALYTICS	PROGRAMME ELECTIVE 3	3	0	0	3

Preamble: This course introduces the salient features and techniques used in soft computing. This course covers the various principles behind processing of image and video, their analytics and their applications. On completion of the course, students will be able to acquire key skills required for solving real-life problems on image and video analytics.

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

CO 1	Interpret the mathematical principles in digital image enhancement and apply them in spatial domain and frequency domain **(Cognitive knowledge level: Apply)**
CO 2	Apply various methods for image filtering and segmentation **(Cognitive knowledge level: Apply)**
CO 3	Make use of various video enhancement and noise reduction techniques **(Cognitive knowledge level: Apply)**
CO 4	Apply various feature extraction and pattern classification techniques on images**.(Cognitive knowledge level: Apply)**
CO 5	Analyse various object detection and recognition techniques on image and video**.(Cognitive knowledge level: Analyse)**
CO 6	Design, develop, implement and present solutions to simple real-life problems with popular open source library using image and video analytical techniques **(Cognitive knowledge level: Apply)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✔		✔	✔		✔	
CO 2	✔		✔	✔		✔	
CO 3	✔		✔	✔		✔	
CO 4	✔		✔	✔	✔	✔	
CO 5	✔		✔	✔	✔	✔	
CO 6	✔	✔	✔	✔	✔	✔	✔

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	40
Analyse	20
Evaluate	Can be evaluated using Course based task/Seminar/Data collection and interpretation/ Assignments
Create	Can be evaluated using Course based task/Seminar/Data collection and interpretation/ Assignments

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.
Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.
There will be two parts; Part A and Part B.
Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks
Total duration of the examination will be 150 minutes.
Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Apply opening and closing operation on the image sample A given below with

structuring element B

A = B = [1 1 1]

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

2. What is Gamma correction and illustrate its need.

3. Verify whether the DFT matrix is unitary or not for N=4.

Course Outcome 2 (CO2)

1. Differentiate Averaging Filter and Weighted Averaging Filter, with an example.

2. Illustrate with an example that the initial threshold in the basic global threshoding algorithm should lie between minimum and maximum values in the image.

3. Illustrate Unsharp Masking and High Boost Filtering.

Course Outcome 3(CO3):

1. For the same line number per frame, what is the relation between the maximum temporalfrequency that a progressive raster can have and that of an interlaced raster which divides each frame into two fields? Also, give the relation between the maximum vertical frequencies.
2. Compare and contrast the features of component and composite format?
3. Illustrate the Video Production, Transmission, and Reception in an Analog Color TV systems, with a block diagram.

Course Outcome 4 (CO4):

1. Illustrate solving the face recognition problem using k-NN classifier.

2. Illustrate the steps in rule based image understanding, with an example.

3. Give case studies on the applications of Content Based Image Retrieval (CBIR).

Course Outcome 5 (CO5):

1. Illustrate texture based object recognition performed?

1. Illustrate the Exhaustive Search Block Matching Algorithm (EBMA) with an example.

Course Outcome 6 (CO6):

1. Prepare a review article based on peer reviewed original publications (referring minimum 10 publications) about the recent image noise filtering techniques.

2. Course project to design and develop a Birds Species Classification System using Support Vector Machine Classifier.

Model Question Paper		
QP CODE:		
Reg No: _______________		
Name: _________________		**PAGES : 4**
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY		
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR		
Course Code: 222ECS013		
Course Name: IMAGE AND VIDEO ANALYTICS		
Max. Marks : 60		**Duration: 2.5 Hours**
PART A		
Answer All Questions. Each Question Carries 5 Marks		
1.	Differentiate Bit-plane slicing and Contrast stretching.	
2.	What is the Convolution property of 2D DFT?	
3.	Differentiate Progressive and Interlaced scan in raster systems.	
4.	What is the role of PCA in a pattern recognition problem?	
5.	Compare pixel based and block based 2D motion representation methods.	**(5x5=25)**
Part B		
(Answer any five questions. Each question carries 7 marks)		

6.	(a)	Find the 4 order Hadamard Transform for the following image segment: $\begin{matrix} 1 & 1 & 2 & 2 \\ 2 & 1 & 1 & 3 \\ 3 & 2 & 1 & 2 \\ 1 & 2 & 2 & 3 \end{matrix}$	**(4)**
	(b)	Demonstrate the steps of generating digital images from the sensed data. Find out the number of bits needed to represent 512 Gray levels.	**(3)**
7.	(a)	Derive two point DFT.	**(4)**
	(b)	Illustrate the significance of negative and logarithmic transformations.	**(3)**
8.	(a)	For the image given below, apply histogram equalization to achieve image enhancement. $f(x,y)= \begin{matrix} 4 & 4 & 4 & 3 & 4 \\ 3 & 4 & 5 & 4 & 3 \\ 3 & 5 & 5 & 5 & 3 \\ 3 & 4 & 5 & 3 & 3 \\ 4 & 4 & 4 & 4 & 4 \end{matrix}$	**(4)**
	(b)	What are the various smoothing non-linear filters in the spatial domain? Which is the best among them and why?	**(3)**

9.	(a)	a) For the image given below, apply Prewitt filter both in X and Y direction and find the resultant image. Also, find the magnitude and angle of gradient at the two underlined pixel positions. $f(x,y)= \begin{matrix} 2 & 0 & 1 & 3 & 2 \\ 1 & \underline{\mathbf{4}} & 7 & 0 & 3 \\ 5 & 4 & \underline{\mathbf{3}} & 5 & 7 \\ 1 & 2 & 0 & 4 & 4 \end{matrix}$	**(4)**
		Illustrate how sharpening can be done in the frequency domain using Butterworth high-pass filters.	**(3)**
10.	(a)	Illustrate the process for forming a composite color video signal. How should you select the color sub-carrier frequency and audio sub-carrier frequency ?	**(4)**
	(b)	What is the perceived color if you mix red, green, and blue dyes in equal proportion ? What is the result if you mix red and green dyes only ?	**(3)**
11.	(a)	Illustrate the steps of Support Vector Machines in solving a two-class object recognition problem.	**(5)**
		Compare the features of SIFT and SURF feature extraction techniques	**(2)**
12.	(a)	Illustrate the main steps in Automatic Traffic Monitoring system using object detection and recognition techniques.	**(4)**
		How is Pixel-Based Motion Estimation performed using Multipoint Neighborhood?	**(3)**

	Syllabus	
Module	**Content**	**Hours**
I	**Fundamentals of Image Processing & Image Enhancement** : Steps in Image Processing Systems, Digital image representation -, Sampling and Quantization, Pixel Relationships, Image Operations – Arithmetic, Geometric & Morphological operations, Colour Models, Image Enhancement in Spatial Domain – Transformations – Negative, Logarithmic, Gamma, Contrast Stretching, Grey level & Bit Plane Slicing. Image Transforms - DFT, DCT, Hadamard Transforms	7
II	**Histogram Processing, Filtering :** Histogram Processing - Histogram Equalisation. Spatial correlation and convolution, Spatial filtering- Smoothing and Sharpening spatial filters, Basics of filtering in frequency domain -Smoothing and sharpening in frequency domain. Image Segmentation- Fundamentals, Thresholding. Edge Detection - Point, Line and Edge Detection , Edge Detection operators.	7
III	**Video Processing :** Video Formation, Perception and Representation: Color Perception and Specification-Human Perception of Color, The Trichromatic Theory of Color Mixture, Color Specification by Tristimulus Values, Color Specification by Luminance and Chrominance Attributes. Video Capture and Display-Principles of Color Video Imaging, Video Cameras, Video Display, Composite versus Component Video, Gamma Correction, Analog Video Raster- Progressive and Interlaced Scan, Characterization of a Video Raster. Analog Color Television Systems- Spatial and Temporal Resolution, Color Coordinate, Signal Bandwidth, Multiplexing of Luminance, Chrominance, and Audio, Analog Video Recording. Digital video- ITU-R BT.601 Digital Video, Other Digital Video Formats and Applications.	8
IV	**Image Analytics :** Feature Extraction - Binary object feature, Histogram based (Statistical) Features, PCA - SIFT – SURF. Visual Pattern Recognition- Patterns and Pattern Classes- Statistical Pattern Classification Techniques- k-Nearest Neighbours Classifier, Support Vector Machines. Image Understanding, Content Based Image	7

	Retrieval.	
V	**Video Analytics & Image and Video Analytics Applications :** Object detection and recognition in video-Texture models, Video classification models- Object tracking in Video – Applications.. Two Dimensional Motion Estimation: Pixel Based Motion Estimation-Block Matching Algorithm. Implementation examples of simple Image and Video processing problems using TensorFlow and Keras. Case Study : Face Detection and Recognition, Automatic Traffic Monitoring.	7

Course Plan

No	Topic	No. of Lectures (36)
1	**Module 1 (Fundamentals of Image Processing & Image Enhancement) 7 Hours**	
1.1	Steps in Image Processing Systems, Digital image representation	1
1.2	Image Operations – Arithmetic, Geometric, Morphological operations	1
1.3	Colour Models	1
1.4	Image Enhancement in Spatial Domain – Transformations – Negative, Logarithmic	1
1.5	Gamma, .Contrast Stretching, Grey level & Bit Plane Slicing	1
1.6	Image Transforms - DFT	1
1.7	DCT, Hadamard Transforms	1
2	**Module 2 (Histogram Processing, Filtering) 7 Hours**	
2.1	Histogram Processing - Histogram Equalisation.	1
2.2	Spatial correlation and convolution,	1
2.3	Spatial filtering- Smoothing, Sharpening spatial filters	1
2.4	Basics of filtering in frequency domain, Smoothing in frequency domain	1
2.5	Sharpening in frequency domain	1
2.6	Image Segmentation- Fundamentals, Thresholding	1
2.7	Edge Detection - Point, Line and Edge Detection, Edge Detection operators.	1
3	**Module 3 (Video Processing) 8 Hours**	
3.1	Video Formation, Perception and Representation: Color Perception and Specification-Human Perception of Color, The Trichromatic Theory of	1

	Color Mixture	
3.2	Color Specification by Tristimulus Values, Color Specification by Luminance and Chrominance Attributes.	1
3.3	Video Capture and Display-Principles of Color Video Imaging, Video Cameras, Video Display	1
3.4	Composite versus Component Video, Gamma Correction,	1
3.5	Analog Video Raster-Progressive and Interlaced Scan, Characterization of a Video Raster.	1
3.6	Analog Color Television Systems- Spatial and Temporal Resolution, Color Coordinate, Signal Bandwidth, Multiplexing of Luminance, Chrominance, and Audio	1
3.7	Analog Video Recording.	1
3.8	Digital video- ITU-R BT.601 Digital Video, Other Digital Video Formats and Applications.	1
4	**Module 4 (Image Analytics) 7 Hours**	
4.1	Feature Extraction , Binary object feature	1
4.2	Histogram based (Statistical) Features, PCA	1
4.3	SIFT, SURF	1
4.4	Visual Pattern Recognition, Patterns and Pattern Classes	1
4.5	Statistical Pattern Classification Techniques- k-Nearest Neighbours Classifier	1
4.6	Support Vector Machines	1
4.7	Image Understanding, Content Based Image Retrieval	1
5	**Module 5 (Video Analytics & Image and Video Analytics Applications) 7 Hours**	
5.1	Object detection and recognition in video, Texture models	1
5.2	Video classification models, Object tracking in Video , Applications	1
5.3	Two Dimensional Motion Estimation	1

5.4	Pixel Based Motion Estimation	1
5.5	Block Matching Algorithm	1
5.6	Implementation examples of simple Image and Video processing problems using TensorFlow and Keras.	1
5.7	Case Study : Face Detection and Recognition, Automatic Traffic Monitoring	1

Reference Books

1. Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", Third Edition, Pearson Education, 2009.
2. Anil K.Jain, "Fundamentals of Digital Image Processing", Pearson Education, 2003
3. Milan Sonka, Vaclav Hlavac and Roger Boyle, "Image Processing, Analysis and Machine Vision", Third Edition, Cengage Learning, 2007.
4. Yao Wang, Jorn Ostermann and Ya-Qin Zhang, "Video Processing and Communications", Prentice Hall, 2001.
5. Simon J. D. Prince, "Computer Vision: Models, Learning, and Inference", Cambridge University Press
6. Caifeng Shan, FatihPorikli, Tao Xiang, Shaogang Gong, "Video Analytics for Business Intelligence", Springer, 2012.
7. Himanshu Singh, "Practical Machine Learning and Image Processing", APress, 2019
8. Vesna Zeljkovic, "Video Surveillance Techniques and Technologies", IGI Global

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS014	INTERNET OF THINGS	PROGRAMME ELECTIVE 3	3	0	0	3

Preamble: This course intends to provide insight into new innovations that will build novel type of interactions among things and humans, and enables the realization of smart cities, infrastructures, and services for enhancing the quality of life and utilization of resources. Knowledge about IOT and its related concepts, different IOT architectures and their components, emerging paradigms such as Fog computing, Platforms and solutions supporting development and deployment of IOT applications, message passing mechanisms such as RPC, REST, and CoAP, data and knowledge management, data confidentiality, data integrity, and operation control issues faced by IOT are included in the course.This course helps the students to develop IOT based applications.

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

CO 1	Make use of the concepts and features of the IOT paradigm.(**Cognitive Knowledge Level: Comprehension)**
CO 2	Analyse Fog computing, TinyOS - nesC and programming frameworks for IOT(**Cognitive Knowledge Level: Analze**)
CO 3	Analyse the data management techniques applied to the IOT environment.(**Cognitive Knowledge Level: Analyze)**
CO 4	Analyse security, and privacy in IOT environments.(**Cognitive Knowledge Level: Apply)**
CO 5	Analyse key enablers and solutions to enable practical IoT systems.(**Cognitive Knowledge Level: Apply)**
CO 6	Design of micro projects for Smart Gas Leakage Detector / Night Patrol at home (**Cognitive Knowledge Level: Apply**)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	√		√			√	
CO 2	√		√	√	√	√	
CO 3	√		√			√	
CO 4	√	√		√		√	
CO 5	√		√		√	√	√
CO 6	√		√			√	

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	50%
Analyse	50%

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

i. Preparing a review article based on peer-reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. "IOT is the framework that enables collaboration between smart mobile devices and cloud." Justify.

2, Explain the state diagram of the open IOT services life cycle with a case study.

Course Outcome 2 (CO2)

1.Justify the need on the four broad requirements that motivate the design of TinyOS.

Course Outcome 3(CO3):

1. Compare Stream Management System (DSMS) and Complex Event Processing (CEP).

Course Outcome 4 (CO4):

1. Illustrate the error detection techniques which are applicable in an IoT environment.

Course Outcome 5 (CO5):

1. What s TWI? How to Configure the TWI for I2C Communication?

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ PAGES : 4

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 222ECS014

Course Name: : INTERNET OF THINGS

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1.	"IOT is the framework that enables collaboration between smart mobile devices and cloud." Justify.	
2.	Explain the major challenges faced in the Fog paradigm.	
3.	How are anomaly detection and categorization of anomalies in data performed?	
4.	Discuss the advantages of obfuscation and diversification techniques.	
5.	What s TWI? How to Configure the TWI for I2C Communication?	**(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.		Explain the state diagram of the open IOT services life cycle.	**(7)**
7.	(a)	Justify the need on the four broad requirements that motivate the design of TinyOS.	**(4)**
	(b)	Explain the design decisions for nesC.	**(3)**
8.		Compare Stream Management System (DSMS) and Complex Event Processing (CEP).	**(7)**
9.		Explain the error detection techniques which are applicable in an IoT environment.	**(7)**

Model Question Paper			
10.		Develop the Station-to-Station protocol (STS) and Identify the two main shortcomings of STS.	**(7)**
11.		Suggest the sensors required to build the environmental-sensing IoT gateway device for weather monitoring.	**(7)**
12.		Develop the steps for the development of a sensor project.	**(7)**

Syllabus and Course Plan

Module 1: **Introduction** (9 Hours) Overview of Internet of Things: Open-source semantic web infrastructure for managing IOT resources in the Cloud - Device/Cloud Collaboration framework for intelligence applications.
Module 2: **Programming frameworks** (11 Hours) Introduction to Fog Computing: principles, architectures, and applications. TinyOS – NesC, Programming frameworks for Internet of Things
Module 3: **Data management techniques** (8 Hours) Stream processing in IoT: foundations, state-of-the-art, and future directions - A framework for distributed data analysis for IoT
Module 4: **Security and privacy** (9 Hours) Security and privacy in the Internet of Things- Internet of Things - robustness and reliability. TinyTO: two-way authentication for constrained devices in the Internet of Things - Obfuscation and diversification for securing the Internet of Things
Module 5: **IoT Implementation** (8 Hours) Creating a simple IoT project - Preparing Raspberry Pi – Interfacing the hardware - Internal representation of sensor values- Persisting data - Creating the actuator project - Creating a controller.

Course Plan

No	Topic	No. of Lectures
1	**Introduction** (7 Hours)	
1.1	Internet of things - definition, evolution. Applications -Smart home applications, Health care, Elder care, Traffic surveillance.	1
1.2	SOA -Based Architecture, API oriented Architecture, Resource Management. Computational Offloading,	1
1.3	Identification and Resource/Service Discovery, IOT Data Management and Analytics, IOT and the CLOUD	1
1.4	Open IOT architecture for IOT/Cloud convergence, Sensor middleware, Cloud computing infrastructure, Directory service, Global Scheduler, Local Scheduler component, Service delivery and utility manager Workflow of open IOT platform	1
1.5	Scheduling process and IOT Services, lifecycle, State diagram of the Open IOT Services lifecycle within the scheduler module. Scheduling and resource management	1
1.6	Resource optimization schemes, Caching technique Service creation flowchart, Comparison of cost - with cache server and public cloud data-score	1
1.7	Runtime adaptation engine, Device/cloud collaboration framework applications of device	1
1.8	cloud collaboration, Semantic QA cache	
2	**Programming frameworks** (9 Hours)	
2.1	Introduction to Fog Computing: principles, architectures, and Applications Motivating scenario for Fog Computing,	1
2.2	Advantages of Fog Computing, Reference architecture of Fog Computing Software-Defined Resource management layer,	1
2.3	Services of Software-Defined Resource management layer, Applications of Fog Computing.	1
2.4	History of TinyOS, Implementation, Requirements motivating the design of TinyOS, Component Model, Interfaces. TinyOS computational concepts Overview of TinyOS Execution Model	1
2.5	Concurrency, TinyOS Theory of Execution: Events & Tasks, TinyOS Architecture. TinyOS-Programming Model.	1
2.6	nesC design, Component Implementation, Design Decisions for nesC, Module Components, Configuration Components Whole-Program Analysis	1
2.7	Detecting Race Conditions, Dealing with Race Conditions, Issues for nesC. Overview of Embedded Programming Languages- nesC, Keil C, Dynamic C, B#	1

2.8	Message Passing in Devices-Remote Procedure Call (RPC), Lightweight RPC (LRPC) Representational state transfer (REST), Computational REST (CREST), Constrained Application Protocol(CoAP), Comparison of HTTP and CoAP	1
2.9	Advantages of CoAP Coordination Languages- Orchestration, Choreography, Linda and eLinda, Orc, Features of Orc, Java Orchestration Language Interpreter Engine (Jolie), Polyglot Programming	1
3	**Data management techniques** (6 Hours)	
3.1	Stream, Stream Processing, Data Stream Management System (DSMS)	1
3.2	differences between two usecases of Stream Processing: Complex Event Processing (CEP)	1
3.3	DSMS and CEP The characteristics of stream data in IOT	1
3.4	general architecture of a stream-processing system in IOT Continuous logic processing system, challenges in stream processing systems. Anomaly detection, problem statement and definitions.	1
3.5	Hyper ellipsoidal anomaly detection.	1
3.6	Distributed anomaly detection.	1
4	**Security and privacy** (8 Hours)	
4.1	IOT security threats, IOT security requirements, security frameworks for IOT, IOT security overview, IOT gateways and security	1
4.2	IOT routing attacks. Security frameworks for IOT - Lightweight cryptography, asymmetric LWC algorithms, privacy in IOT networks	1
4.3	IOT characteristics and reliability issues, reliability challenges.	1
4.4	Addressing reliability, security aspects and solutions	1
4.5	TinyTO: Two-way authentication for constrained devices in the Internet of Things.	1
4.6	TinyTO protocol. BCK with pre-shared keys for TinyTO, handshake implementation	1
4.4	IOT network stack and access protocols, Obfuscation and diversification techniques .	1
4.5	Enhancing the security in IOT using obfuscation and diversification techniques, motivations and limitations, different use-case scenarios on software diversification and obfuscation	1
5	**IoT Implementation** (7 Hours)	
5.1	Three key components to an IOT architecture, Sensor to gateway communication	1
5.2	wired gateway interfaces, wireless gateway interfaces. Sensors - sensors required to build the environmental	1
5.3	sensing IOT gateway device for weather monitoring. Gateway,	1

	Gateway hardware, Gateway software.	
5.4	Data transmission - advanced message queuing protocol, backend processing, to CLOUD or not to cloud	1
5.5	Creating a simple sensor project - Preparing Raspberry Pi – Clayster libraries. Hardware, Interfacing the hardware	1
5.6	Internal representation of sensor values- Persisting data. External representation of sensor values, Exporting sensor data.	1

Text Book

1. RajkumarBuyya; Amir VahidDastjerdi , "Internet of Things", Morgan Kaufmann, 2016

Reference Books

1. Peter Waher, "Learning Internet of Things", Packt Publishing, 2015

2. S. SitharamaIyengar; Nandan Parameswaran; Vir V. Phoha; N. Balakrishnan; Chuka Okoye, "Fundamentals of Sensor Network Programming: Applications and Technology", Wiley, December 14, 2010

3. Robert Stackowiak, Art Licht, VenuMantha, Louis Nagode, "Big Data and The Internet of Things: Enterprise Information Architecture for A New Age", Apress, 2015

Web Resources

1. https://www.coursera.org/specializations/internet-of-things

2. http://web.mit.edu/professional/digital-programs/courses/IoT

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS015	STATISTICAL NLP	PROGRAMME ELECTIVE 3	3	0	0	3

Preamble: This course helps to familiarize with the basics of statistical techniques for processing human languages and large text corpora. It covers Word Sense Disambiguation, Part-of-Speech Tagging, Machine Translation etc.At the end of the course, the learner will be able to develop natural language-based applications and frameworks.

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

CO 1	Apply the concepts of Probability theory and linguistic essentials for processing large text corpora. **(Cognitive Knowledge Level: Apply)**
CO 2	Apply the concepts of Word Sense Disambiguation for processing human languages **(Cognitive Knowledge Level: Apply)**
CO 3	Apply the concepts of Part of speech tagging and Probabilistic Context free Grammarfor processing natural languages **(Cognitive Knowledge Level: Apply)**
CO 4	Analyse the concepts of Machine Translation for human languages **(Cognitive Knowledge Level: Analyse)**
CO5	Apply the concepts of Clustering and Text Categorization for natural language processing to solve real-life problems. **(Cognitive Knowledge Level: Apply)**

Program Outcomes (PO)

Students will be able to demonstrate the following attributes after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1				✓	✓	✓	
CO 2	✓			✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	✓
CO 5	✓			✓	✓	✓	

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	40
Analyse	20
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred): 15 marks

ii. Course-based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course-based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. It is determined that in a text 25 of every 100 verbs and 600 of every 1,000 nouns are inflected. If the number of nouns on a particular page is four times more than that of verbs, calculate the probability of

   a. A word without inflection being randomly selected.

   b, A noun with inflection being randomly selected.

   C. A verb without inflection is randomly selected.

2. Analyse the different issues involved in processing a corpus? Describe the techniques to resolve them.

Course Outcome 2 (CO2)

1. Construct phrase structure grammar of the following sentences.

   a) I prefer morning flight.

   b) Show me the last flight to leave.

2. Explain flip flop algorithm for disambiguation and implement the partition based on the given data.

   Translation{t1,...,tm}: {take, make, rise, speak}

   Indicators{x1,....,xn}: {mesure, note, exemple, decision, parole}

Course Outcome 3(CO3):

1. Is it possible to use HMM for POS tagging? Justify your answer.

2. Illustrate the tagging and partial parsing in question- answering system with appropriate examples.

Course Outcome 4 (CO4):

1. Describe the noisy channel model in machine translation.

2. Compare and contrast different machine translation strategies.

Course Outcome 5 (CO5):

1. When is the maximum entropy framework appropriate as a classification method?

2. Illustrate how the Vector Space model helps in ad-hoc retrieval?

Model Question Paper :COMPUTER SCIENCE AND ENGINEERING-CS2			
QP CODE:			
Reg No: _______________			
Name: _________________ **PAGES: 4**			
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY			
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR			
Course Code:			
Course Name:			
Max. Marks : 60 **Duration: 2.5 Hours**			
PART A			
Answer All Questions. Each Question Carries 5 Marks			
1.	Discuss the problem of word sense disambiguation as a noisy channel model		
2.	Explain Witten-Bell smoothing.		
3.	Illustrate the usage of Penn Treebank in Probabilistic Parsing?		
4.	Compare syntactic and semantic transfer approaches for machine translation.		
5.	Describe the advantages and disadvantages of Latent Semantic Indexing.		**(5x5=25)**
Part B			
(Answer any five questions. Each question carries 7 marks)			
6.	(a)	A text contains 10 nouns and 5 verbs. A word is selected and replaced by the other category and then a second word is selected. Calculate : 1. The probability that the second word is a verb. 2. The probability that both words selected belong to the same category. 3. The probability that both words selected are of different category.	**(4)**
	(b)	Compute the t-value of words *new companies* in a corpus with following counts C(new)=30,000, C(companies)=9,000, C(new companies)=20 and corpus size N=15,000,000	**(3)**
7.	(a)	What are the limitations of Maximum Likelihood Estimation (MLE)? Give solutions to overcome these limitations.	**(4)**
	(b)	Explain Dictionary-Based Disambiguation with example	**(3)**

8. Consider a simple three-state Markov model of the weather. Any given day, the weather can be described as being precipitation (rain or snow), cloudy or sunny. Given that the weather on day t=1 is sunny, what is the probability that the weather for the next 7 days will be "sun, sun, rain, rain, sun, clouds, sun" ? **(7)**

9. Find the probability of the string "Astronomers saw stars with ears" with the help of inside probability for the given PCFG grammar. **(7)**

S→ NP VP (1.0)	P → with (1.0)
NP → NP PP (0.4)	V → saw (1.0)
PP → P NP (1.0)	NP → ears (0.18)
VP → V NP (0.7)	NP → astronomers (0.1)
VP → VP PP (0.3)	NP → stars (0.18)
NP→ saw (0.04)	NP→telescopes (0.1)

10. Measuring length in number of characters is preferable because the variance in number of words is greater. Do you agree that word-based length is more variable? Why? **(7)**

11. (a) Explain the following: **(3)**

a) Perceptrons

b) KNN classification

(b) We have a document collection D with 5 documents. Given a query q, we know that three documents are relevant to q. A retrieval algorithm produces the ranking (of all documents in D) shown in the following table. **(4)**

Rank i	+/-		p(i)	r(i)	F-Score
1	+				
2	+				
3	-				
4	+				
5	-				

i) Calculate precision, recall and F-Score for each document.

ii) Calculate average precision.

iii) Is there any breakeven point?

12.		Let the random variable X has three possible outcomes {a, b, c}. Consider two distributions on this random variable.	(7)

Symbol	p(x)	q(x)
a	1/2	1/3
b	1/4	1/3
c	1/4	1/3

a) Calculate H(p), H(q)

b) Verify that in this case D(p \|\| q) != D(q \|\| p)

Syllabus

	Syllabus	
Module	**Content**	**Hours**
I	Introduction-Mathematical Foundations: Elementary Probability theory- Probability spaces, Conditional probability and independence, Bayes' theorem, Random variables, Joint and conditional distributions, Bayesian statistics, Information Theory- Entropy, Joint entropy and conditional entropy, The noisy channel model, Relative entropy or Kullback-Leibler divergence, The relation to language: Cross entropy, Linguistic Essentials- Parts of Speech and Morphology, Phrase Structure, Semantics and Pragmatics, Corpus-Based Work- Corpora, tokenization, marked-up data, Zipf's laws, Collocations, Concordances	8
II	Statistical Inference: n-gram Models over Sparse Data, n-gram, Building n-gram models, Statistical Estimators, Combining Estimators, Word Sense Disambiguation - Supervised Disambiguation, Dictionary-Based Disambiguation, Lexical Acquisition	8

III	Grammar-Markov Models, Hidden Markov Models-Implementations, properties and variants, Part-of-Speech Tagging-Information Sources in Tagging, Markov Model Taggers, Hidden Markov Model Taggers, Transformation-Based Learning of Tags, Probabilistic Context free Grammar-features, Probability of a String, Problems with the Inside-Outside Algorithm	9
IV	Probabilistic parsing-Parsing for disambiguation, Parsing models vs. language models, Tree probabilities and derivational probabilities, Phrase structure grammars and dependency grammars, Statistical Alignment -Text alignment, word alignment, Statistical Machine Translation	8
V	Clustering-Hierarchical Clustering, Non-Hierarchical Clustering, Text **Categorization-Decision** Trees, Maximum Entropy Modelling, Perceptrons, k- Nearest Neighbour Classification, Topics in Information Retrieval	7

Course Plan

No	Topic	No. of Lectures (40)
1	**Module 1 (Linguistic Essentials)**	
1.1	Introduction, Mathematical Foundations: Elementary Probability theory-Probability spaces, Conditional probability and independence	1
1.2	Bayes' theorem, Random variables, Joint and conditional distributions, Bayesian statistics	1
1.3	Information Theory- Entropy, Joint entropy and conditional entropy, The noisy channel model	1
1.4	Relative entropy or Kullback-Leibler divergence, The relation to language: Cross entropy	1
1.5	Linguistic Essentials- Parts of Speech and Morphology,	1
1.6	Phrase Structure, Semantics and Pragmatics	1
1.7	Corpus-Based Work- C o r p o r a, tokenization, marked-up data	1
1.8	Zipf's laws, Collocations, Concordances	1
2	**Module 2 (Word Sense Disambiguation)**	
2.1	Statistical Inference: n-gram Models over Sparse Data, n-gram	1
2.2	Building n-gram models	1
2.3	Statistical Estimators	1
2.4	Combining Estimators	1
2.5	Word Sense Disambiguation	1

2.6	Supervised Disambiguation	1
2.7	Dictionary-Based Disambiguation	1
2.8	Lexical Acquisition	1
3	**Module 3 (Part-of-Speech Tagging)**	
3.1	Grammar-Markov Models	1
3.2	Hidden Markov Models- Implementations, properties and variants,	1
3.3	Part-of-Speech Tagging- Information Sources in Tagging	1
3.4	Markov Model Taggers	1
3.5	Hidden Markov Model Taggers	1
3.6	Transformation-Based Learning of Tags	1
3.7	Probabilistic Context free Grammar -features	1
3.8	Probability of a String	1
3.9	Problems with the Inside-Outside Algorithm	1
4	**Module 4 (Statistical Alignment and Machine Translation)**	
4.1	Probabilistic parsing	1
4.2	Parsing for disambiguation	1
4.3	Parsing models vs. language models	1
4.4	Tree probabilities and derivational probabilities	1
4.5	Phrase structure grammars and dependency grammars	1
4.6	Statistical Alignment -Text alignment	1
4.7	Word alignment	1
4.8	Statistical Machine Translation	1
5	**Module 5 (Clustering and Text Categorization)**	
5.1	Clustering-Hierarchical Clustering	1
5.2	Non-Hierarchical Clustering	1
5.3	Text Categorization-Decision Trees	1
5.4	Maximum Entropy Modelling	1
5.5	Perceptrons	1
5.6	k- Nearest Neighbour Classification	1
5.7	Topics in Information Retrieval	1

Reference Books

1. C.D. Manning and H. Schubert: Foundations of Statistical Natural Language Processing: MIT Press, 2001

2. Jurafsky, D. and J. H. Martin, Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall, 2000.

3. Charniak, E.: Statistical Language Learning. The MIT Press

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS016	KNOWLEDGE BASED SYSTEM DESIGN	PROGRAMME ELECTIVE 3	3	0	0	3

Preamble: This course helps to familiarize with the architecture and design principles of knowledge-based systems. It covers Knowledge Acquisition, Knowledge Representation and the development of Knowledge-Based Systems. On successful completion of the course, the students will be able to distinguish between human and artificial experts, their problem-solving strategies, and their knowledge representation methods.

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

CO 1	Understand the fundamentals of knowledge engineering **(Cognitive Knowledge Level: Understand)**
CO 2	Apply the concepts of problem solving for building Expert System **(Cognitive Knowledge Level: Apply)**
CO 3	Apply the concepts of Knowledge based system for developing applications**.** **(Cognitive Knowledge Level: Apply)**
CO 4	Apply knowledge representation methods for developing Knowledge based system **(Cognitive Knowledge Level: Apply)**
CO 5	Make use of Semantics of Expert Systems for building applications. **(Cognitive Knowledge Level: Apply)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓						
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓		✓		
CO 4			✓	✓			
CO 5			✓			✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	40
Analyse	20
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Describe the role of inference engine in Rule-based expert system.
2. Illustrate the difference between an expert system and a knowledge based system?

Course Outcome 2 (CO2)

1. Explain conflict resolution technique in rule based systems.
2. Compare Hayes-Roth's andClancey's classification of expert system

Course Outcome 3(CO3):

1. Illustrate the advantages and limitations of a Knowledge based system.

2.. Describe the components of knowledge base system and illustrate the need of each component with an wexample.

Course Outcome 4 (CO4):

1. Discuss knowledge acquisition strategies in MUD system

2. Differentiate between procedural knowledge and factual knowledge

Course Outcome 5 (CO5):

1. What makes case-based reasoning different from rule based reasoning? Which one will be better for knowledge based system design?

2. What are the sources of uncertain reasoning?

Model Question Paper :COMPUTER SCIENCE AND ENGINEERING-CS2		
QP CODE:		
Reg No: _______________		
Name: _________________ **PAGES : 4**		
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY		
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR		
Course Code:		
Course Name:		
Max. Marks : 60 **Duration: 2.5 Hours**		
Answer All Questions. Each Question Carries 5 Marks		
1	How does an expert system help in solving complex problems?	
2	Why don't rule based systems backtrack much?	
3	Why don't all expert systems use Bayes' Rule to compute the degrees of certainty associated with conclusions? Is there any other method to solve such problems?	
4	Why is Knowledge acquisition a bottle neck? What solutions have been proposed?	
5	What is a case library and how does a case differ from a book?	**(5x5=25)**

Part B
(Answer any five questions. Each question carries 7 marks)

6	(a)	What are the characteristics and components of an expert system?	**(3)**
	(b)	Differentiate 'search space' and 'solution space'. What will be the search space and solution space for the game of chess?	**(4)**

7	(a)	How can we select a knowledge engineering tool? How hard such tools are to learn and use?	**(4)**
	(b)	List out the classification of expert system tasks.	**(3)**
8.		Describe the Knowledge-Based System architecture and its working.	(7)
9.	(a)	Identify some knowledge acquisition strategies for heuristics classification.	(3)
	(b)	What all difficulties may arise in the implementation phase of building a Knowledge-Based System?	(4)
10.		When the search space is potentially large, which type of testing is beneficial? Explain how it works?	(7)
11.	(a)	Describe the working of any industrial or commercial expert systems.	(4)
	(b)	What do you mean by Knowledge acquisition? What are the stages of knowledge acquisition?	(3)
12.		What are the common pitfalls during Knowledge Based Systems implementation and how we can avoid them?	(7)

Syllabus

	Syllabus	
Module	**Content**	**Hours**
I	Introduction to Knowledge Engineering: The Human Expert and an Artificial Expert –Knowledge based systems architecture-Types of Knowledge- Characteristics of Knowledge-Components of Knowledge-Basic structure of knowledge based system- Knowledge Base -Inference Engine	9
II	Problem Solving Process- Rule Based Systems – Heuristic Classifications – Constructive Problem Solving – Toolsfor Building Expert Systems - Expert System Architectures	6
III	Knowledge Based Systems: Objectives of KBS - Components of KBS - Categories of KBS – Difficulties- KBS Architecture: Basic Structure- Applications	7

IV	Developing Knowledge-Based Systems: Difficulties - Development Model - Knowledge Acquisition - Developing Relationships with Experts - Sharing Knowledge - Dealing with Multiple Experts - Knowledge Representation-Factual Knowledge - Procedural Knowledge, Knowledge Management	10
V	Case Based Reasoning – Semantics of Expert Systems – Modelling of Uncertain Reasoning. Machine Learning – Rule Generation and Refinement – Learning Evaluation – Testing and Tuning.	7

Course Plan

No	Topic	No. of Lectures (39)
1	**Module 1 (Knowledge Engineering)**	
1.1	Introduction to Knowledge Engineering	1
1.2	The Human Expert and an Artificial Expert	1
1.3	Knowledge based systems architecture	1
1.4	Types of Knowledge	1
1.5	Characteristics of Knowledge	1
1.6	Components of Knowledge	1
1.7	Basic Structure of Knowledge-Based Systems	1
1.8	Knowledge Base	1
1.9	Inference Engine	1
2	**Module 2 (Building Expert Systems)**	
2.1	Problem Solving Process	1
2.2	Rule Based Systems	1
2.3	Heuristic Classifications	1
2.4	Constructive Problem Solving	1
2.5	Tools for Building Expert Systems	1
2.6	Expert System Architectures	1
3	**Module 3 (Knowledge Based Systems)**	
3.1	Knowledge Based Systems	1
3.2	Objectives	1
3.3	Components of Knowledge based system	1
3.4	Categories of Knowledge based system	1
3.5	Difficulties	1
3.6	Basic Structure of KBS	1
3.7	Applications of KBS	1
4	**Module 4 (Development of Knowledge-Based Systems)**	
4.1	Developing Knowledge-Based Systems: Difficulties	1

4.2	Development Model	1
4.3	Knowledge Acquisition	1
4.4	Developing Relationships with Experts	1
4.5	Sharing Knowledge	1
4.6	Dealing with Multiple Experts	1
4.7	Knowledge Representation	1
4.8	Factual Knowledge	1
4.9	Representing Procedural Knowledge	1
4.10	Knowledge management	1
5	**Module 5 (Case Based Reasoning)**	
5.1	Case Based Reasoning	1
5.2	Semantics of Expert Systems	1
5.3	Modelling of Uncertain Reasoning.	1
5.4	Machine Learning	1
5.5.	Rule Generation	1
5.6	Refinement Learning Evaluation	1
5.7	Testing and Tuning.	1

Reference Books

1. Peter Jackson. Introduction to Expert Systems. 3 rd Edition, Pearson Education 2007.

2. Rajendra Arvind Akerkar . Knowledge-Based Systems. Jones and Bartlet Publishers 2010

3. Robert I. Levine, Diane E. Drang, Barry Edelson. AI and Expert Systems: a comprehensive guide, C language, 2nd edition, McGraw-Hill

4. Jean-Louis Ermine: Expert Systems: Theory and Practice. 4th printing, Prentice-Hall ofIndia, 2001

5. Stuart Russell, Peter Norvig: Artificial Intelligence: A Modern Approach. 2nd Edition

6. N.P.Padhy: Artificial Intelligence and Intelligent Systems. 4th impression , Oxford University Press, 2007.

SEMESTER II

PROGRAM ELECTIVE IV

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS017	SOFT COMPUTING	PROGRAMME ELECTIVE 4	3	0	0	3

Preamble: This course introduces the salient features and techniques used in soft computing. This course covers various neural networks models, fuzzy systems, genetic algorithm concepts and their applications. On completion of the course, students will be able to acquire key skills required for solving real-life problems using soft computing techniques.

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

CO 1	Analyse soft computing techniques and their applications (**Cognitive knowledge level: Analyze)**
CO 2	Make use of various neural network models in problem solving (**Cognitive knowledge level: Apply)**
CO 3	Examine various fuzzy concepts (**Cognitive knowledge level: Analyse)**
CO 4	Illustrate use of fuzzy models in various applications **(Cognitive knowledge level: Apply)**
CO 5	Analyse genetic algorithms and their applications **(Cognitive knowledge level: Analyze)**
CO 6	Design, develop, implement and present solutions to simple real-life problems with popular open source library using soft computing techniques **(Cognitive knowledge level: Apply)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1			✔	✔		✔	
CO 2	✔		✔	✔	✔	✔	
CO 3			✔	✔		✔	
CO 4	✔		✔	✔		✔	
CO 5	✔		✔	✔		✔	
CO 6	✔	✔	✔	✔	✔	✔	✔

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	60
Analyse	40
Evaluate	Can be evaluated using Assignments/ projects
Create	Can be evaluated using Assignments/ projects

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Compare linear and non-linear activation functions with an example for each.

2. Show that a single layer perceptron cannot simulate a simple exclusive-OR function and give the reason.

3. Implement logical NOR function using neural network model

Course Outcome 2 (CO2)

1.Design a suitable neural network for human face recognition.

2. Illustrate with an example how Supervised Learning technique can be used to solve a multilabel classification problem

3. Compare sigmoidal and step activation functions.

Course Outcome 3(CO3):

1. Two fuzzy sets A and B defined on the English alphabets (F,E,X,Y,I,T) are defined as below:

A={(F,0.2), (E,0.9), (X,0.4), (Y,0.3), (I,0.8), (T,0.5)}

B={(F,0.4), (E,0.6), (X,0.7), (Y,0.2), (I,0.8), (T,0.6)}

Find the following.

i) A ∩ B   ii) A U Bc   iii) Verify Demorgan's Law (A U B)c =Ac∩ Bc

2. Consider a set P={P1,P2,P3,P4} of four varieties of paddy plants,set D={D1,D2,D3,D4} of the various diseases affecting the plants and S={S1,S2,S3,S4} be the common symptoms of the diseases.Let R be a relation on P X D and S be a relation on D X S.

$$R = \begin{array}{c|cccc} & D1 & D2 & D3 & D4 \\ \hline P1 & 0.4 & 0.2 & 0.9 & 0.5 \\ P2 & 0.8 & 0.2 & 0.2 & 0.6 \\ P3 & 0.6 & 0.6 & 0.7 & 0.9 \\ P4 & 0.3 & 0.5 & 0.4 & 0.4 \end{array} \qquad S = \begin{array}{c|cccc} & S1 & S2 & S3 & S4 \\ \hline D1 & 0.2 & 0.2 & 0.7 & 0.9 \\ D2 & 1 & 0 & 0.4 & 0.5 \\ D3 & 0.3 & 1 & 0.5 & 0.8 \\ D4 & 0.6 & 0 & 0.7 & 0.4 \end{array}$$

Obtain the association of the plants with the different symptoms of the diseases using max – min composition.

3. Fuzzy binary relation R is defined on set A={1,2,3, 4, 5} and B ={90, 91, 92, 93, 94, 95 } and represents the relation "a is much smaller than b" . It is denoted by the membership function

$\Bigg\{$

R(a,b)= $1 - \frac{a}{b}$ for a ≤ b

0 Otherwise

Where a € A and b € B. Find the domain and range of R.

Course Outcome 4 (CO4):

1. Design a fuzzy controller to determine the wash time of a washing machine. Assume the input is dirt and grease. Use three descriptors for input variable and five descriptors for output variable. Derive the set of rules for controller action and defuzzification. Show that if the dirt are more , the wash time will be more and vice versa.

2. Illustrate the multiatribute decision making with an example.

3. Demonstrate problem solving using Mamdani fuzzy model, with an example.

Course Outcome 5 (CO5):

1. Illustrate Roulette Wheel Selection technique with an example.

2. Differentiate single-point crossover, multipoint crossover and uniform crossover techniques used in Genetic Algorithm.

Course Outcome 6 (CO6):

1. Prepare a review article based on peer reviewed original publications (referring minimum 10 publications) about recent neuro-fuzzy systems based applications.

2. Course project to design and develop handwritten character recognition system using Python/ TensorFlow/ Keras based on MNIST handwritten digit dataset.

3. Present a seminar on any one application of Genetic Algorithm (Other than Traveling Salesman Problem)

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES : 3**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 222ECS017

Course Name: Soft Computing

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Implement the following truth table using suitable neural network model.

Input		Output
X1	X2	
0	0	0
0	1	1
1	0	1
1	1	1

2. Differentiate Supervised and Unsupervised Learning Neural Networks with an example for each.

3. Differentiate Generalised Bell membership function and Left-Right Membership Function.

4. Discuss Fuzzy Expert System with an example

5. Explain the steps in Genetic Algorithm, with an example.

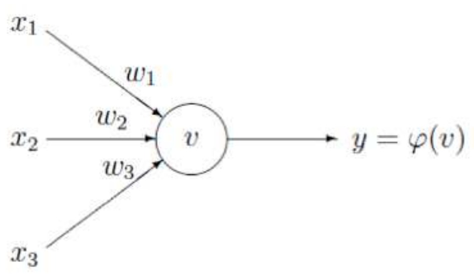
(5x5=25)

Part B

(Answer any five questions. Each question carries 7 marks)

6.	(a)	A handwritten character recognition system for English language has to be designed. Use feedforward neural network system to design the same.	**(4)**
	(b)	Implement logical X-OR function using neural network model	(3)
7.	(a)	A company has collected large amount of data in pairs of input output vectors. Design a system using radial basis function network for predicting output for new input.	**(4)**

(b) Consider the unit shown in the following figure. (3)



Suppose that the weights corresponding to the three inputs have the following values: w1= 1, w2= **- 3**, w3= 2 and the activation of the unit is given by the step-function:

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Calculate what will be the output value y of the unit for each of the following input patterns:

Pattern	P_1	P_2	P_3	P_4
x_1	1	0	1	1
x_2	0	1	0	1
x_3	0	1	1	1

8. (a) Let P and Q be two fuzzy relations defined on the sets X x Y and Y x Z respectively,where X={x1, x2}, Y={y1, y2, y3,}, and Z={z1, z2}. P and Q are given using the following relation matrices. **(4)**

P = [0.50.20.30.10.70.6] Q = [0.500.80.40.61]

Find the association P o Q using both max-min composition and max-product compositions

(b) Differentiate Triangular and trapezoidal fuzzy membership functions. (3)

<table>
<tr><td>9.</td><td>(a)</td><td>Consider a two input – one output problem that includes following three rules.

Rule: 1 - IF x is A3 OR y is B1 THEN z is C1

Rule: 2 - IF x is A2 AND y is B2 THEN z is C2

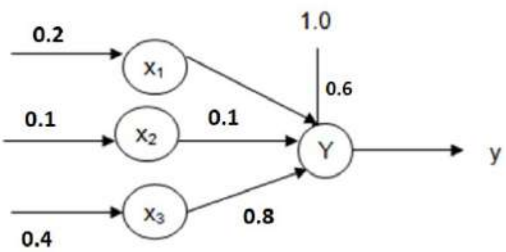
Rule: 3 - IF x is A1 THEN z is C3

Assume A3=0.2, B1=0.4, A2=0.7,B2=0.9,A1=0.3 and the selected values of z in c1=(0,11,21), C2=(,40,50,60, 70) & C3=(60, 70,80,90).Find out the value of z using Mamdani fuzzy inference system.</td><td>(5)</td></tr>
<tr><td></td><td>(b)</td><td>Derive the conclusion for Multiple Rules with multiple antecedents.</td><td>(2)</td></tr>
<tr><td>10.</td><td>(a)</td><td>Illustrate the problem-solving method using Sugeno fuzzy system, with an example.</td><td>(4)</td></tr>
<tr><td></td><td>(b)</td><td>A fuzzy reasoning system is provided with the following facts and rules.

)Premise 1 (fact):x is A’ and y is B’

)Premise 2 (Rule 1): if x is A and y is B , then z is C.

Demonstrate the inference procedure to find out the conclusion z is C’</td><td>(3)</td></tr>
<tr><td>11.</td><td>(a)</td><td>Calculate the output of the neuron y for the following network using Binary Sigmoidal activation function.


</td><td>(4)</td></tr>
<tr><td></td><td>(b)</td><td>Implement logical AND function using neural networks.</td><td>(3)</td></tr>
<tr><td>12.</td><td>(a)</td><td>A salesman has to follow the shortest route to visit N cities exactly once and reach the starting city. Apply genetic algorithm to solve this problem.</td><td>(4)</td></tr>
<tr><td></td><td>(b)</td><td>Illustrate Tournament Selection in Genetic Algorithm with an example.</td><td>(3)</td></tr>
</table>

Syllabus

Module 1 - Introduction

Soft Computing Constituents, Computational Intelligence and Soft Computing vs Artificial Intelligence and Hard Computing. Introduction to artificial neural networks- biological neurons, Mc Culloch and Pitts models of neurons, Perception networks, Multilayer Perceptron. Types of activation function, network architectures, learning process, Learning XOR and other logical Gate functions, Sigmoid Neurons, Gradient Descent Algorithm,Back propagation Neural Network.

Module 2 – Neural Networks

Adaptive Networks, Feed Forward Networks, Supervised Learning Neural Networks, Radial Basis Function Networks, Reinforcement Learning, Unsupervised Learning Neural Networks, Adaptive Resonance Architectures, BAM, Maxnet, Kohonen Self Organizing Maps, K-means clustering algorithm, Introduction to Convolutional Neural Networks, Applications of ANNs to solve real life problems. Familiarisation of neural network tools , Building neural networks with Python, TensorFlow, Introduction to Keras.

Module 3 – Fuzzy Sets & Logic

Fuzzy versus Crisp, Fuzzy sets, Membership function, linguistic variable, basic operators, properties, Extension principle and Fuzzy relations, Cartesian product, Operations on Fuzzy sets , Operations on fuzzy relations, Fuzzy Logic

Module 4 – Fuzzy Reasoning

Fuzzy If-Then Rules, Fuzzy Inference Systems, Fuzzy Expert Systems, Fuzzification and Defuzzification methods, Fuzzy Decision Making , Mamdani Fuzzy Models, Sugeno Fuzzy Models, Applications of Fuzzy logic, Neuro Fuzzy Systems.

Module 5 – Genetic Algorithm

Introduction to genetic algorithm and hybrid systems - Genetic algorithms , natural evolution , properties, classification, GA features, coding, selection, selection methods, reproduction , cross over and mutation operators, basic GA and structure, Solving Travelling Salesman Problem using GA. Applications using GA, Hybrid Systems - GA based BPNN.

Course Plan

No	Topic	No. of Lectures (36 Hours)
1	**Module 1 (Introduction to Soft Computing And Neural Networks) -7 Hours**	
1.1	Soft Computing Constituents, Computational Intelligence and Soft Computing vs Artificial Intelligence and Hard Computing. Introduction to artificial neural networks, biological neurons	1
1.2	Mc Culloch and Pitts models of neurons, Perception networks	1
1.3	Multilayer Perceptron, Types of activation function	1
1.4	Network architectures, Learning process	1
1.5	Learning XOR and other logical Gate functions	1
1.6	Sigmoid Neurons, Gradient Descent Algorithm	1
1.7	Back propagation Neural Network.	1
2	**Module 2 (Neural Networks) 8 Hours**	
2.1	Adaptive Networks, Feed Forward Networks, Supervised Learning Neural Networks	1
2.2	Radial Basis Function Networks, Reinforcement Learning,	1
2.3	Unsupervised Learning Neural Networks, Adaptive Resonance Architectures,	1
2.4	BAM, Maxnet, Kohonen Self Organizing Maps,	1
2.5	K-means clustering algorithm, Introduction to Convolutional Neural Networks,	1
2.6	Applications of ANNs to solve real life problems. Familiarisation of neural network tools	1
2.7	Building neural networks with Python	1
2.8	TensorFlow, Introduction to Keras.	1

3	**Module 3 (Fuzzy Sets & Logic) 7 Hours**	
3.1	Fuzzy versus Crisp, Fuzzy sets, Membership function, linguistic variable	1
3.2	Basic operators, properties	1
3.3	Fuzzy relations	1
3.4	Cartesian product	1
3.5	Operations on Fuzzy sets	1
3.6	Operations on fuzzy relations	1
3.7	Fuzzy Logic	1
4	**Module 4 (Fuzzy Reasoning) 7 Hours**	
4.1	Fuzzy If-Then Rules	1
4.2	Fuzzy Inference Systems, Fuzzy Expert Systems	1
4.3	Fuzzification and Defuzzification methods	1
4.4	Fuzzy Decision Making	1
4.5	Mamdani Fuzzy Models	1
4.6	Sugeno Fuzzy Models	1
4.7	Applications of Fuzzy logic, Neuro Fuzzy Systems.	1
5	**Module 5 (Genetic Algorithm) 7 Hours**	
5.1	Introduction to genetic algorithm and hybrid systems	1
5.2	Genetic algorithms , Natural evolution , Properties	1
5.3	Classification, GA features, Coding, Selection, Selection methods	1
5.4	Reproduction - Cross over and Mutation operators	1
5.5	Basic GA and structure, Solving Travelling Salesman Problem using GA	1
5.6	Applications using GA	1
5.7	Hybrid Systems - GA based BPNN	1

Reference Books

1. Jyh-Shing Roger Jang, Chuen-Tsai Sun, Eiji Mizutani, Neuro-Fuzzy and Soft Computing, Prentice-Hall of India, 2003.
2. S.N. Sivanandam, S.N. Deepa, “Principles of Soft Computing”, 2/e, John Wiley India, 2012.
3. James A. Freeman and David M. Skapura, Neural Networks Algorithms, Applications, and Programming Techniques, Pearson Edn., 2003.
4. Simon Haykin "Neural Networks and Learning Machines" Prentice Hall, 2008
5. Jon Krohn, Grant Beyleveld, Aglae Bassens, Deep Learning Illustrated, Pearson, 2020
6. Timothy J. Ross, “Fuzzy Logic with engineering applications”,Wiley India, 2021
7. Bart Kosko.: Neural Network and Fuzzy Systems-Prentice Hall, Inc., Englewood Cliffs, 1991
8. Goldberg D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning-Addison Wesley, 1989
9. Mitchell Melanie, An Introduction to Genetic Algorithm, Prentice Hall, 1998.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS018	HIGH PERFORMANCE COMPUTING	PROGRAMME ELECTIVE 4	3	0	0	3

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.

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

CO1	Analyse different types of modern processing environments and parallel computing hardware **(Cognitive Knowledge Level: Analyse)**
CO2	Summarize the concepts of Instruction Level Parallelism **(Cognitive Knowledge Level: Analyse)**
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	Comprehend the advanced features of GPU architecture**. (Cognitive Knowledge Level: Analyse)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓			✓	
CO 2	✓		✓			✓	
CO 3	✓		✓			✓	
CO 4	✓		✓			✓	
CO 5	✓		✓			✓	

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	40
Analyse	20
Evaluate	-
Create	-

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.
There will be two parts; Part A and Part B.
Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks
Total duration of the examination will be 150 minutes.
Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

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 : 2**
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY			
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR			
Course Code: 222ECS018			
Course Name: High Performance Computing			
Max. Marks : 60			**Duration: 2.5 Hours**
PART A			
Answer All Questions. Each Question Carries 5 Marks			
1.	Discuss in detail the importance of considering processor performance for the design of an efficient computer system.		
2.	Illustrate how data-level parallelism is achieved in vector, and SIMD architectures.		

3.	Consider the following loop:
	`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.		

4.	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?		
5.	Explain about GPU thread engine.		**(5x5=25)**
Part B **(Answer any five questions. Each question carries 7 marks)**			
6.	(a)	Illustrate how processes are protected with the help of virtual memory.	**(4)**
	(b)	Describe the quantitative principle of computer design with Amdahl's law.	**(3)**
7.	(a)	Explain in detail data dependence and hazards.	**(7)**

Model Question Paper			
	(b)		
8.	(a)	Describe the unique features of very long instruction word processors.	**(3)**
	(b)	Consider a three-way superscalar machine renaming these three instructions concurrently: `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?	**4**

9.	(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.
		`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 */`
		`}`
			(7)

10	(a)	Demonstrate the different types of hardware approaches required for the working of multithreading.	**(7)**
11	(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?	**(7)**
12	(a)	Explain the multi-GPU platform.	**(7)**

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

Course 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

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.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS019	CRYPTOGRAPHY AND NETWORK SECURITY	PROGRAMME ELECTIVE 4	3	0	0	3

Preamble: The course introduces the fundamental concepts of cryptography and the security issues. It covers basic symmetric and asymmetric cryptographic algorithms, cryptographic hash functions, message authentication protocols, digital signatures, key management and distribution schemes for symmetric and asymmetric encryption, and case studies involving different types of attacks and conventional algorithms. The course enables learners to apply the cryptographic concepts to real-life problems.

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

CO 1	Summarize basic cryptographic algorithms and security issues. (**Cognitive Knowledge Level: Apply**)
CO 2	Compare and analyse various symmetric and asymmetric key cryptographic algorithms. (**Cognitive Knowledge Level: Apply**)
CO 3	Demonstrate cryptographic hash functions and digital signature schemes. (**Cognitive Knowledge Level: Apply**)
CO 4	Summarize key management and distribution schemes for symmetric and asymmetric encryption. (**Cognitive Knowledge Level: Apply**)
CO 5	Apply the concepts to design an efficient cryptographic algorithm, hash function and digital signature scheme. (**Cognitive Knowledge Level: Apply**)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams.

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards.

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects.

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓	✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	
CO5	✓		✓	✓		✓	

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	60
Analyse	40
Evaluate	-
Create	-

Assignments or course projects can be used for higher level assessment of course outcomes.

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks iii. Test

paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.
There will be two parts; Part A and Part B.
Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks. Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Use Vigenere Cipher with key HEALTH to encrypt the message “Cryptography and network security”.

2. In a public key system using RSA, you intercept the cipher text C=8 sent to a user whose public key is e=13, n=33. What is the plain text M?

3. Alice wants to send a message to Bob, without Eve observing it. Alice and Bob have agreed to use a symmetric cipher. Key exchange has already occurred, and so they share a key K. Outline the steps that Alice and Bob must follow when they encrypt and decrypt, respectively.

Course Outcome 2 (CO2)

1. With the help of an example, illustrate the concept behind ElGamal Crypto

System. 2. Elaborate on an application which uses Chinese Reminder Theorem.

3. Consider a Diffie Hellman scheme with a common prime q = 11 and primitive root α = 2.
   i. Show that 2 is a primitive root of 11.

   ii. If user A has public key $Y_A = 9$, what is A's private key?

   iii. If user B has public key $Y_B = 3$, what is the shared secret key K, shared with

Course Outcome 3 (CO3):

1. Discuss on an attack to which MAC is vulnerable. How can MAC be made more secure?

2. Imagine that a specific hash function is used for HMAC, and that vulnerabilities have been found in the hash functions so that the HMAC is insecure. Which changes in the HMAC are required in order to make it secure again?

3. The Diffie-Hellman key agreement algorithm achieves key agreement by allowing two hosts to create a shared secret.

   a. Clearly explain the operation of the Diffie–Hellman key exchange protocol.

   b. Clearly explain why the basic Diffie–Hellman protocol does not provide any assurance regarding which other party the protocol is run with.

Course Outcome 4 (CO4):

1. Explain the authentication procedures defined by X.509 certificate.

2. Illustrate the concept of 'certificate chain' for verification of digital signature on X.509 certificate.

3. Describe the steps in finding the message digest using SHA-512 algorithm. Analyse the situation where two messages have the same message digest.

Course Outcome 5 (CO4):

1. Prepare a report on various symmetric and asymmetric key cryptosystems. Compare features of both the systems based on their applications.

2. Present a seminar on different message authentication protocols and digital signature

schemes elaborating on their unique features.

3. Course project to design a private or public cryptosystem/ message authentication protocol or digital signature scheme.

Model Question Paper		
QP CODE:		
Reg No: _______________		
Name: _________________ PAGES : 4		
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY		
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR		
Course Code: 222ECS019		
Course Name: R for Data Science		
Max. Marks : 60 Duration: 2.5 Hours		
PART A		
Answer All Questions. Each Question Carries 5 Marks		
1.	Encrypt the message “the movie is set to release this Friday” using Vigenere cipher with key “awards”. Ignore the space between words. Decrypt the message to get the plain text.	
2.	Alice and Bob agreed to use RSA algorithm for the secret communication. Alice securely choose two primes, p=5 and q=11 and a secret key d=7. Find the corresponding public key. Bob uses this public key and sends a cipher text 18 to Alice. Find the plain text.	
3.	Compare digital signatures with authentication protocols.	

4.		Explain why message authentication alone is insufficient as proof of message origin in general, and to settle disputes about whether messages have been sent.	
5.		Quoting suitable examples, differentiate between IP spoofing attacks and denial of service attacks.	**(5x5=25)**
		Part B (Answer any five questions. Each question carries 7 marks)	
6.		The encryption key in a transposition cipher is (3,1,4,5,2). Perform encryption and decryption for the message "meet me after the toga party". Add a bogus character at the end to make the last group the same size as the others.	**(7)**
7.		Elaborate on the algorithm for generating keys in RSA algorithm. Perform encryption and decryption using RSA for the following. P=7; q=11; e=13; M=8.	**(7)**

8.		Illustrate man in the middle attack on Diffie Hellman key exchange algorithm.	**(7)**
9.		Alice wants to send a message M with a digital signature Sig(M) to Bob. Alice and Bob have an authentic copy of each other's public keys, and have agreed on using a specific hash function h. Outline the steps that Alice must follow when signing M, and the steps that recipient Bob must follow for validating the signature Sig(M).	**(7)**
10.		Alice wants to send a message to Bob. Alice wants Bob to be able to ensure that the message did not change in transit. Briefly outline the cryptographic steps that Alice and Bob must follow to ensure the integrity of the message by creating and verifying a MAC.	**(7)**
11.		Explain Kerberos authentication mechanism with suitable diagram.	**(7)**
12.		With the help of an example, illustrate IP spoofing attacks.	**(7)**

Syllabus

	Syllabus	
Module	**Content**	**Hours**
I	**Security Concepts -** Introduction, The need for security, Security approaches, Principles of security, Types of Security attacks. Services and Mechanism, A model for Network Security. **Cryptography Concepts and Techniques**: Introduction, plain text and cipher text, substitution techniques, transposition techniques, encryption and decryption, symmetric and asymmetric key cryptography, steganography, key range and key size, possible types of attacks.	**9**
II	**Symmetric Key Cryptography** - Block Cipher principles, DES, AES, Blowfish, RC5, IDEA, Block cipher operation, Stream ciphers, RC4. **Asymmetric Key Cryptography -** Principles of public key cryptosystems, RSA algorithm, Elgamal Cryptography, Diffie-Hellman Key Exchange, Knapsack Algorithm.	**11**
III	**Cryptographic Hash Functions** - Message Authentication, Secure Hash Algorithm (SHA-512), Message authentication codes: Authentication requirements, HMAC, CMAC, Digital signatures, Elgamal Digital Signature Scheme.	**7**
IV	**Key Management and Distribution** - Symmetric Key Distribution Using Symmetric& Asymmetric Encryption, Distribution of Public Keys, Kerberos, X.509 Authentication Service, Public – Key Infrastructure.	**6**
V	**Case Studies of Cryptography** - Denial of service attacks, IP spoofing attacks, Secure inter branch payment transactions, Conventional Encryption and Message Confidentiality, Conventional Encryption Principles, Conventional Encryption Algorithms, Location of Encryption Devices, Key Distribution.	**7**

Course Plan

No	Topic	No. of Lectures
1	**Security and Cryptography Concepts (9 hours)**	
1.1	**Security Concepts -** Introduction, The need for security, Security approaches	1
1.2	Principles of security, Types of Security attacks.	1
1.3	Services and Mechanism	1
1.4	A model for Network Security.	1
1.5	**Cryptography Concepts and Techniques**: Introduction, plain text and cipher text, substitution techniques	1
1.6	Transposition techniques	1
1.7	Encryption and decryption	1
1.8	Symmetric and asymmetric key cryptography, steganography	1
1.9	Key range and key size, possible types of attacks.	1
2	**Symmetric and Asymmetric Key Cryptography (11 hours)**	
2.1	**Symmetric Key Cryptography** - Block Cipher principles, DES	1
2.2	AES	1
2.3	Blowfish	1
2.4	RC5	1
2.5	IDEA	1
2.6	Block cipher operation	1
2.7	Stream ciphers, RC4	1
2.8	**Asymmetric Key Cryptography -** Principles of public key cryptosystems, RSA algorithm	1
2.9	Elgamal Cryptography	1
2.10	Diffie-Hellman Key Exchange	1
2.11	Knapsack Algorithm.	1
3	**Cryptographic Hash Functions (7 hours)**	

3.1	**Cryptographic Hash Functions** - Message Authentication	1
3.2	Secure Hash Algorithm (SHA-512)	1
3.3	Message authentication codes: Authentication requirements	1
3.4	HMAC	1
3.5	CMAC	1
3.6	Digital signatures	1
3.7	Elgamal Digital Signature Scheme.	1
4	**Key Management and Distribution (6 hours)**	
4.1	Symmetric Key Distribution Using Symmetric & Asymmetric Encryption	1
4.2	Symmetric Key Distribution Using Symmetric &Asymmetric Encryption	1
4.3	Distribution of Public Keys	1
4.4	Kerberos	1
4.5	X.509 Authentication Service	1
4.6	Public – Key Infrastructure.	1
5	**Case Studies of Cryptography (7 hours)**	
5.1	Denial of service attacks, IP spoofing attacks	1
5.2	Secure inter branch payment transactions	1
5.3	Conventional Encryption and Message Confidentiality	1
5.4	Conventional Encryption Principles	1
5.5	Conventional Encryption Algorithms	1
5.6	Conventional Encryption Algorithms	1
5.7	Location of Encryption Devices, Key Distribution.	1

References:

1. William Stallings, Cryptography and Network Security –6th Edition, Pearson Education, 2013.

2. Behrouz A. Forouzan, Debdeep Mukhopadhyay, Cryptography and Network Security, Special Indian Edition, Mc Graw Hill Education, 2007.

3. Cryptography and Network Security : Atul Kahate, Mc Graw Hill, 2nd Edition, 2008.

4. Cryptography and Network Security: C K Shyamala, N Harini, Dr T R Padmanabhan, Wiley India, 1st Edition, 2011.

5. Information Security, Principles, and Practice: Mark Stamp, Wiley India, 2011. 6. Principles of Computer Security: WM. Arthur Conklin, Greg White, TMH, 2018.

7. Introduction to Network Security: Neal Krawetz, Cengage Learning, 2007.

8. Network Security and Cryptography: Bernard Menezes, Cengage Learning, 2010.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS020	**BIG DATA ANALYTICS ON GENOMIC DATA**	**PROGRAMME ELECTIVE 4**	**3**	**0**	**0**	**3**

Preamble: This course helps the learners to provide practical/research solutions to problems in the domain of Bioinformatics. It enables the learners to understand. concepts of Bioinformatics, Application of AI in Bioinformtics, Big Data Bioinformatics and Data Analytics with NGS data. This course helps the learners to develop practical solutions to problems in bioinformatics.

Course Outcomes: The COs shown are only indicative. For each course, there can be 4 to 6 COs.

After the completion of the course the student will be able to

CO 1	Analyse basic concepts of Biomolecules, Biological databases, Sequence characteristics **(Cognitive Knowledge Level :Apply)**
CO 2	Apply ML/DL Model for RNA /Protein structure Analysis **(Cognitive Knowledge Level :Apply)**
CO 3	Apply Big data techniques in Bioinformatics **(Cognitive Knowledge Level :Apply)**
CO 4	Comprehend the Data Analytics pipelines for NGS data **(Cognitive Knowledge Level :Apply)**
CO 5	Design and Develop RNASeq /Chip/ Metagenomics seq Pipelines **(Cognitive Knowledge Level :Apply)**
CO 6	Develop/suggest a solution for any research problems in the field of Bioinformatics **(Cognitive Knowledge Level :Apply)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✔		✔	✔		✔	
CO 2	✔		✔	✔		✔	
CO 3	✔		✔	✔		✔	
CO 4	✔		✔	✔	✔	✔	
CO 5	✔			✔	✔	✔	
CO 6	✔	✔	✔	✔	✔	✔	✔

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	50-80%
Analyse	20-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.
There will be two parts; Part A and Part B.
Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks
Total duration of the examination will be 150 minutes.
Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Comprehend the nature and scope of Bioinformatics
2. Articulate the different biomolecules and the various databases
3. Comprehend the concept of sequence Alignments

Course Outcome 2 (CO2)

1 Comprehend and apply ML/DL models for Bioinformatics data

2. Apply LSTM /GAN models
3. Comprehend Transformer based architecture

Course Outcome 3(CO3):

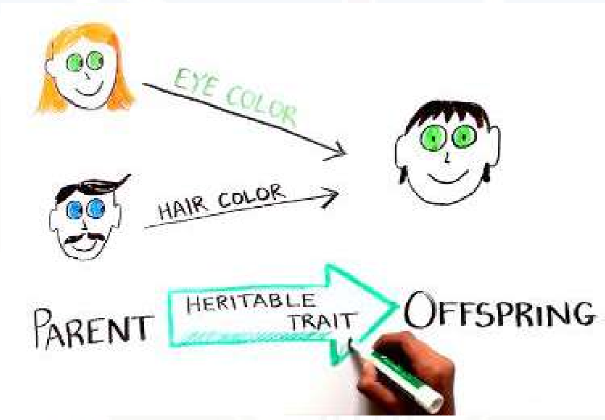
1. Comprehend the Big data concept, challenges and the various techniques in it.
2. Apply and Analyse Map Reduce concept with Hadoop and Spark
3. Apply Spark SQL for scaling genomic data

Course Outcome 4 (CO4):

1. Comprehend the understanding of NGS pipelines
2. Articulate on Denovo assemblies

Course Outcome 5 (CO5):

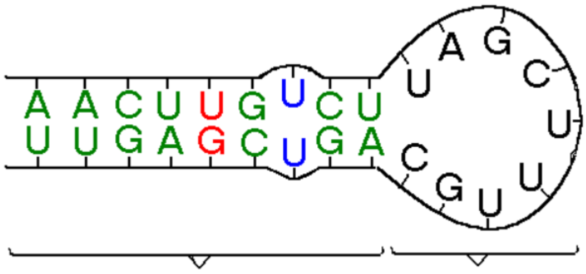
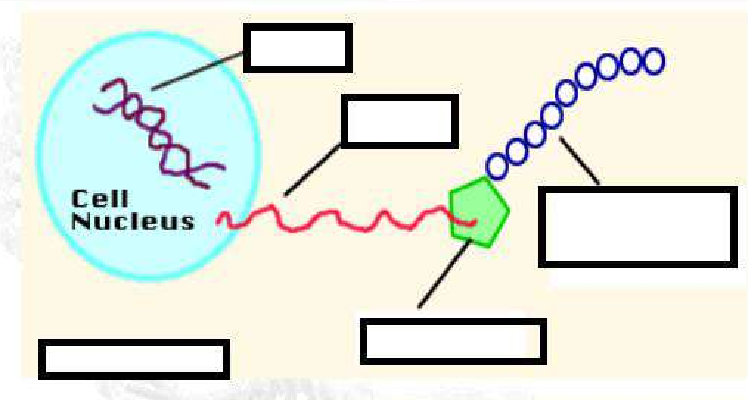
1. Demonstrate Understanding of RNASeq /Chipseq Pipelines
2. Apply the principle of QIIME

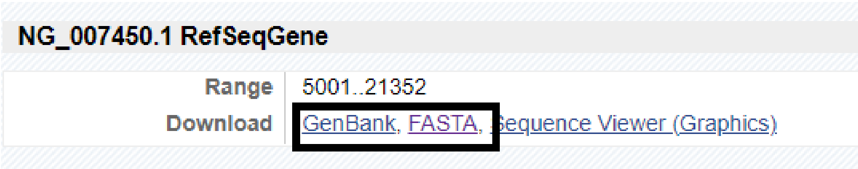
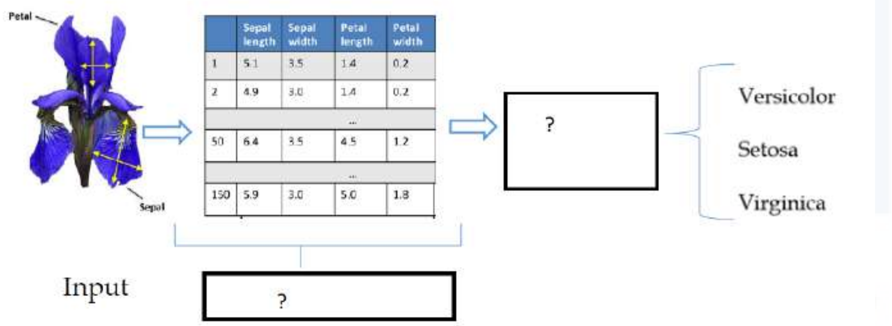
Model Question Paper		
QP CODE:		
Reg No: _______________		
Name: _________________ **PAGES : 4**		
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY		
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR		
Course Code: 222ECS020		
Course Name: BIG DATA ANALYTICS ON GENOMIC DATA		
Max. Marks : 60 **Duration: 2.5 Hours**		
PART A		
Answer All Questions. Each Question Carries 5 Marks		
1.	(i) Study the cartoon and comment on the heritable trait that transmitted from parent to offspring  (ii) A structural gene has two DNA strands and for the partial sequence of the gene given below, which is the template strand and why? 5'ACTCGATGCTAG3' 3'TGAGCTACGATC5'	
2.	Bioinformatics is the application of computational technology to handle the rapidly growing repository of information related to molecular biology". Give a brief insight based on your knowledge on this definition of bioinformatics	

3.	Write short note on the need and application of Biological databases in sequence analysis		
4.	Write a Python program using Pyspark to perform the following task. • Create a dataframe with the following attributes student id, name, Marks • Find the scholars whose mark <50% using spark sql comment • Find the average and variance of the marks • Find the maximum and minimum marks		
5.	Discuss the steps involved in the Metagenomics data analysis pipeline.		**(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	(a)	For the given RNA molecule secondary structure, identify the (i) Stem (ii) Loop and explain the importance of stem-loop structure of RNA. 	**3 marks**
	(b)	The protein synthesising process the shown below. Identify the different stages and explain the role of mRNA?  	**4 marks**
7.	(a)	How do you submit biological data to a public database? List the major submission tools in NCBI.	**3 marks**

	(b)	From NCBI, while downloading a gene sequence, you can download the sequence in two formats- GenBank & FASTA. To know more information about the sequence, which format will be useful? Justify your answer.  	**4 marks**
8.	(a)	Align AGCTCAG with AGGTCA by using a suitable scoring scheme	**5 marks**
	(b)	Differentiate between Pairwise and Multiple sequence alignment.	**2 marks**
9.	(a)	Write a simple python program for illustrating the map reduce model used in Hadoop?	**4 marks**
		Explain with examples the PageRank algorithm using Map Reduce programming concept	**3 marks**
10	(a)	Study the pictorial representation given below. Comment on the unknown block and its relevance in machine learning  	**3 marks**
	(b)	Explain architecture of CNN. Estimate the number of parameters using 2 convolution layer model	**4 marks**
11	(a)	Briefly explain the big data challenges in the field of Bioinformatics.	**3 marks**
	(b)	Distinguish between Bowtie and BWA.	**4 marks**
12	(a)	What is a contig? What is a scaffold? Describe de novo assembly. Name a software used for de novo assembly.	**3 marks**
	(b)	Briefly explain the public sequence databases which support Next generation sequencing data and the various data formats being supported	**4 marks**

Syllabus

Module 1: Introduction to Bioinformatics

Informational view of life science; Introduction to Biomolecules through games-Foldit (Protein overlapping), Eterna Game (RNA Structure), History of Bioinformatics, Biological databases, Sequence Similarity, identity and homology, Scoring Matices, Sequence Alignment, Phylogeny

Module 2: AI in Bioinformatics

AI applications in the field of genomics, the role of deep learning and data mining in computational biology and bioinformatics, ML/DL algorithms: ANN, CNN, LSTM, BERT, GAN, Protein structure prediction (deep neural networks), RNA structure prediction (deep learning models), RNA-protein binding sites prediction with CNN, Deep neural net to predict target gene expression, Transcription Factor Binding via MLP, LSTM, CNN, Protein Contextual Embeddings via BERT

Module 3 Big Data Bioinformatics

Overview of Big data , Data Storage and Analysis, Processing-SCV principle, Batch Vs Stream processing, Big data Analytics- Typical Analytical Architecture -Types (Descriptive, Inquisitive, Predictive, Perspective); Visualisation and Applications, Computational facilities for analysing Big data – Cluster computing vs. Cloud computing, – Challenges in Big Data Analytics , Big data Frameworks - MapReduce, Hadoop and Spark, Spark SQL and dataframes, Spark for Bioinformatics, Big data analytics using Python- PySpark, Big data processing for DNA sequence analysis – PASTASpark

Module 4 Data Analytics with NGS data

Introduction to next generation sequencing: NGS Platforms, NGS, advantages, limitations and applications, NGS Data sources: NCBI SRA, EBI-ENA, DDBJ-SRA; SRA toolkit; NGS Data analysis: FASTQ files, Quality check, Pre-processing, Mapping - Principles, tools - BWA, Bowtie, SAM tools -output file formats -BAM, SAM; Denovo assemblies - Principles, tools - SOAPdenovo, Velvet; Visualization tools - IGV; Whole Genome/Exome pipeline for Variant calling - VCF files**.**

Module 5 Advanced Data Analytics with NGS data

RNAseq - Gene expression analysis, Differential expression analysis. Alternative splicing - TopHat and Cufflinks for RNAseq, ChIPseq - Introduction and biological theories on ChIPseq analysis. DNA fragment evaluation. Peak identification. Two condition comparison. Saturation analysis. Motif finding and related theories, Metagenomics analysis using QIIME and Picrust

Course Plan (For 3 credit courses, the content can be for 40 hrs and for 2 credit courses, the content can be for 26 hrs. The audit course in third semester can have content for 30 hours).

No	Topic	No. of Lectures (40 hrs)
1	**Module 1 : Introduction to Bioinformatics**	
1.1	Informational view of life science; Definition; DNA-RNA and Protein as information, Primary and secondary structure of DNA,	1
1.2	Chargaff's Rules, Different forms of DNA, RNA, Introduction to Biomolecules through games-Foldit (Protein overlapping), Eterna Game (RNA Structure)	1
1.3	History of Bioinformatics, Definition of Bioinformatics, Bioinformatics versus Computational Biology, Goals of Bioinformatics analysis,	1
1.4	Biological data bases :- File format, conversion of file format, Data retrieval system, Genome browsers. Biological data file formats	1
1.5	Basic concepts of sequence similarity, identity and homology, Scoring matrices- PAM and BLOSUM matrices,	
1.6	Data retrieval system, Sequence databases-EMBL, GenBank, DDBJ;	
1.7	Protein databases- UniProt, Protein Data Bank.	
1.8	Concept of sequence alignment- pairwise and multiple Pairwise- Local and global, ,	1
1.9	Dot plot, BLAST, Multiple sequence alignment (MSA) – CLUSTAL Omega	
1.10	Phylogeny: Basic concepts of phylogeny, Phylogenetic tree construction using MEGA.	1
2	**Module 2 : AI in Bioinformatics**	
2.1	AI applications in the field of genomics, the role of deep learning and data mining in computational biology and bioinformatics.	1
2.2	ML/DL algorithms: ANN, CNN, LSTM,	1
2.3	ML/DL algorithms: BERT, GAN	1
2.4	ML model for protein expression /Sequence classification,	1
2.5	Protein structure prediction (deep neural networks), RNA structure prediction (deep learning models),	1
2.6	RNA-protein binding sites prediction with CNN, Deep neural net to predict target gene expression	1
2.7	Transcription Factor Binding via MLP, LSTM, CNN	1
2.8	Protein Contextual Embeddings via BERT	1
3	**Module 3 Big Data Bioinformatics**	
3.1	Overview of Big data – Definition, Characteristics, Sources, Types-	1

	Structured, Unstructured & Semi-structured;	
3.2	Data Storage and Analysis -NAS, DAS, NoSQL databases; Processing-SCV principle, Batch Vs Stream processing	1
3.3	Big data Analytics- Typical Analytical Architecture – Requirement for new analytical architecture -Types (Descriptive, Inquisitive, Predictive, Perspective); Visualisation and Applications,	1
3.4	Computational facilities for analysing Big data – Cluster computing vs. Cloud computing, – Challenges in Big Data Analytics –Need of big data frameworks	1
3.5	Big data Frameworks - MapReduce, Hadoop and Spark, Spark SQL and dataframes,	1
3.6	Spark for Bioinformatics, Big data analytics using Python- PySpark,	1
3.7	Big data processing for DNA sequence analysis – PASTASpark	1
4	**Module 4 Data Analytics with NGS data**	
4.1	Introduction to next generation sequencing: NGS Platforms,	1
4.2	NGS technologies (WGS, ChIP-seq & RNA-seq), advantages, limitations and applications.	1
4.3	NGS Data sources: NCBI SRA, EBI-ENA, DDBJ-SRA;	1
4.4	SRA toolkit; NGS Data analysis: FASTQ files, Quality check, Pre-processing,	1
4.4	Mapping - Principles, tools - BWA, Bowtie, SAM tools -output file formats -BAM, SAM;	1
4.5	Denovo assemblies - Principles, tools -	1
4.6	SOAPdenovo, Velvet; Visualization tools - IGV;	1
4.7	Whole Genome/Exome pipeline for Variant calling - VCF files.	1
5	**Module 5 Advanced Data Analytics with NGS data**	
5.1	RNAseq - Gene expression analysis,.	1
5.2	Differential expression analysis. Alternative splicing -	1
5.3	TopHat and Cufflinks for RNAseq	1
5.4	ChIPseq - Introduction and biological theories on ChIPseq analysis.	1
5.5	DNA fragment evaluation. Peak identification. Two condition comparison. Saturation analysis.	1
5.6	Motif finding and related theories	1
5.7	Metagenomics analysis using QIIME and Picrust	1

Reference Books

- Brown, T.A. 2002 Genome. John Wiley Press, US.
- Campbell, A.M. & Heyer, L.J. 2002 Discovering Genomics, Proteomics and Bioinformatics. Benjamin/Cummings
- Stuart M. Brown. Next-Generation DNA Sequencing Informatics, Second Edition. New York University School of Medicine (ISBN-13: 978-1621921236)
- Xinkun Wang. Next Generation Sequencing Data Analysis, CRC Press. (ISBN13: 9781482217889).
- Mueller J. P. & Massaron L. (2016),"Machine learning for dummies",John Wiley & Sons.
- Russell S. J. & Norvig P. (2003), "Artificial Intelligence A Modern Approach", Pearson Education.
- Jordan M. I. & Mitchell T. M. (2015), "Machine learning: Trends, Perspectives and Prospects. Science", 349(6245), pp. 255-260. (Journal Article)
- Kamath, U., Liu, J., & Whitaker, J. (2019). *Deep learning for NLP and speech recognition* (Vol. 84). Cham: Springer.
- Gulli, A., Kapoor, A., & Pal, S. (2019). *Deep learning with TensorFlow 2 and Keras: regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API*. Packt Publishing Ltd.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS021	PATTERN RECOGNITION	PROGRAMME ELECTIVE 4	3	0	0	3

Preamble: This course aims to impart the fundamentals of statistical pattern recognition and neural network techniques. It introduces to the learner the various pattern recognition algorithms, feature selection, classification, clustering and the use of neural networks in feature extraction. This helps the learner to apply the algorithms in applications that works on pattern recognition and machine intelligence.

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

CO 1	Apply probability and numerical methods in statistical pattern recognition. **(Cognitive Knowledge Level: Apply)**
CO 2	Apply statistical methods in feature selection.**(Cognitive Knowledge Level: Apply)**
CO 3	Apply linear algebra and statistical methods in parameter and non-parameter estimation.**(Cognitive Knowledge Level: Apply)**
CO 4	Apply the technique of decision trees in pattern recognition. **(Cognitive Knowledge Level: Apply)**
CO 5	Analyze the use of deep learning networks andartificial neural networks in pattern recognition. **(Cognitive Knowledge Level: Analyze)**
CO 6	Design, Develop, Implement and Present innovative ideas in problem solving with various pattern recognition techniques. **(Cognitive Knowledge Level: Create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and developmentwork in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓	✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓		✓	✓	✓	✓	
CO 4	✓		✓	✓	✓	✓	
CO 5	✓		✓	✓	✓	✓	
CO 6	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%-80%
Analyze	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii.Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.
There will be two parts; Part A and Part B.
Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks
Total duration of the examination will be 150 minutes.
Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Illustrate the design cycle of a pattern recognition system with the help of an example.
2. Suppose that we have three coloured bottles r (red), b(blue) and g(green).Box r contains 3 apples, 4 oranges and 3 limes. Box B contains 1 apple, 1 orange and 0 limes and box g contains 3 apples, 3 oranges and 4 limes. If a box is chosen at random with probability p(g)=0.2, p(b)=0.2 and p(g)=0.6 and piece of fruit is removed from the box(with equal probability of selecting items from the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

Course Outcome 2 (CO2):

1. Illustrate feature selection using t-Test with the help of an example.

Course Outcome 3(CO3):

1. Derive the fuzzy C spherical shells (FCSS) algorithm for the case that spherical clusters are to be identified.

Course Outcome 4 (CO4):

1. Illustrate decision tree with the help of an example. How does it enable pattern classification?

2. Construct a decision tree using the following data.

Outlook	**Temp**	**Humidity**	**Windy**	**Play Golf**
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Cloudy	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Cloudy	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Cool	Normal	True	Yes
Cloudy	Mild	High	True	Yes
Cloudy	Mild	Normal	False	Yes
Sunny	Hot	High	True	No

Course Outcome 5 (CO5):

1. How do artificial neural networks play a significant role in pattern recognition? Also discuss about its parameter optimisation techniques.

Course Outcome 6 (CO6):

1. Suppose an accident prone area is under surveillance and real time CCTV visuals are available to you. Design a solution to automatically detect accidents on the road from

those real time CCTV visuals. Explain about any one pattern recognition algorithm you will make use here and how?

Model Question Paper

QP CODE:

Reg No: _______________

Name: _________________ **PAGES: 2**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 222ECS021

Course Name: Pattern Recognition

Max. Marks : 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1.	In a town it was estimated that 3% of people have a particular disease. Adiagnosis test was conducted for all the people, which yielded 8% false positiveand 92% true positive results. A person is found as positive after the test. What isthe probability that this person is truly having the disease?	**(5)**
2.	How does morphological operations play a role in pattern recognition?	**(5)**
3.	How can visual imagery be analysed using convolutional neural networks?	**(5)**
4.	How does a decision tree handle continuous attributes?	**(5)**
5.	Define the terms: weights, bias, activations with respect to neural networks	**(5)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	Illustrate the design principles of pattern recognition system with an example.	**(7)**
7.	Derive the fuzzy C spherical shells (FCSS) algorithm for the case that spherical clusters are to be identified.	**(7)**

8. Show that in the case of Gaussian distributions the Chernoff bound becomes **(7)**

$$\epsilon_{CB} = \exp(-b(s))$$

where

$$b(s) = \frac{s(1-s)}{2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T [s\Sigma_j + (1-s)\Sigma_i]^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \frac{1}{2}\ln\frac{|s\Sigma_j + (1-s)\Sigma_i|}{|\Sigma_j|^s |\Sigma_i|^{1-s}}$$

Then take the derivative with respect to s and show that for equal covariance matrices the optimum is achieved for $s = 1/2$. Thus, in this case $b(s)$ equals the Bhattacharyya distance.

9. Let N_1, N_2 be the available values of a feature in two classes, respectively. The feature is assumed to follow a Gaussian distribution with the same variance in each class. Define the test statistic **(7)**

$$q = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{s_z\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

where

$$s_z^2 = \frac{1}{N_1 + N_2 - 2}\left(\sum_{i=1}^{N_1}(x_i - \bar{x})^2 + \sum_{i=1}^{N_2}(y_i - \bar{y})^2\right)$$

and μ_1, μ_2 are the respective true mean values. Show that q follows the t-distribution with $N_1 + N_2 - 2$ degrees of freedom.

10. Discuss the significance of pre-processing in feature selection. Illustrate any two methods used for pre-processing. **(7)**

11. How can artificial neural networks be applied in Pattern recognition? Also illustrate the features of recurrent neural networks. **(7)**

12. Construct a decision tree using the following data. **(7)**

Outlook	**Temp**	**Humidity**	**Windy**	**Play Golf**
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Cloudy	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Cloudy	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Cool	Normal	True	Yes
Cloudy	Mild	High	True	Yes
Cloudy	Mild	Normal	False	Yes
Sunny	Hot	High	True	No

Syllabus

Module 1: Introduction to Pattern Recognition

Basics of pattern recognition systems, various applications, Machine Perception, classification of pattern recognition systems.Design of Pattern recognition system, Pattern recognition Life Cycle. Statistical Pattern Recognition: Review of probability theory, Gaussian distribution. Normal density and discriminant functions.

Module 2: Feature Selection

Feature selection – Outlier removal – Data normalization – Missing data, The Peaking phenomenon, Feature selection using statistical hypothesis testing- Hypothesis testing basics – Application of t-Test in feature selection. Class separability measures-Divergence-Chernoff bound and Bhattacharya distance-Scatter matrices, Feature subset selection –Scalar feature selection, Feature vector selection.

Module 3: Clustering Algorithms

Unsupervised learning and clustering - Criterion functions for clustering. Cluster validation.
Fuzzy clustering algorithms- Point representatives- quadratic surfaces and representatives – hyper plane representatives. Binary morphology clustering algorithms (BMCAs) – Discretization – Morphological operations - Determination of clusters in a discrete binary set-

Assignment of feature vectors to clusters – The algorithmic scheme, Boundary detection algorithms.

Module 4: Dimensionality reduction

Dimensionality reduction: Principal component analysis - its relationship to Eigen analysis. Fisher discriminant analysis - Generalised Eigen analysis. Eigen vectors/Singular vectors as dictionaries. Factor Analysis, Total variability space - a dictionary learning method.Non negative matrix factorisation - a dictionary learning method.
Linear discriminant functions: Gradient descent procedures, Perceptron.

Module 5:Artificial neural networks and Pattern Classification

Artificial neural networks: Review of Artificial neural network concepts, convolutional neural networks, recurrent neural networks.
Non-metric methods for pattern classification: Non-numeric data or nominal data. Decision trees: Classification and Regression Trees (CART).

Course Plan

No.	Topic	No. of Lectures (40 Hours)
1	**Module 1: Introduction to Pattern Recognition**	**7**
1.1	Basics of pattern recognition systems, applications	1
1.2	Machine Perception, Classification of pattern recognition systems	1
1.3	Design of Pattern recognition system	1
1.4	Pattern recognition Life Cycle	1
1.5	Statistical Pattern Recognition	1
1.6	Review of probability theory	1
1.7	Normal density and discriminant functions	1
2	**Module 2: Feature Selection**	**10**
2.1	Feature selection – Outlier removal	1
2.2	Data normalization – Missing data	1
2.3	The peaking phenomenon	1
2.4	Feature selection using statistical hypothesis testing	1
2.5	Hypothesis testing basics – Application of tTest in feature selection	1
2.6	Class separability measures-Divergence	1
2.7	Chernoff bound and Bhattacharya distance	1
2.8	Scatter matrices	1
2.9	Feature subset selection –Scalar feature selection	1
2.10	Feature vector selection	1
3	**Module 3: Clustering Algorithms**	**9**

3.1	Unsupervised learning and clustering	1
3.2	Criterion functions for clustering. Cluster validation.	1
3.3	Fuzzy clustering algorithms- Point representatives	1
3.4	Quadratic surfaces and representatives – hyper plane representatives.	1
3.5	Binary morphology clustering algorithms (BMCAs)	1
3.6	Discretization	1
3.7	Morphological operations - Determination of clusters in a discrete binary set	1
3.8	Assignment of feature vectors to clusters	1
3.9	The algorithmic scheme, Boundary detection algorithms.	1
4	**Module 4: Dimensionality reduction**	**8**
4.1	Principal component analysis - its relationship to Eigen analysis	1
4.2	Fisher discriminant analysis	1
4.3	Generalised Eigen analysis	1
4.4	Eigen vectors/Singular vectors as dictionaries	1
4.5	Total variability space - a dictionary learning method	1
4.6	Non negative matrix factorisation - a dictionary learning method	1
4.7	Linear discriminant functions: Gradient descent procedures	1
4.8	Perceptron	1
5	**Module 5: Artificial neural networks and Pattern Classification**	**6**
5.1	Review of Artificial neural networks, Introduction to deep neural networks	1
5.2	Convolutional neural networks	1
5.3	Recurrent neural networks	1
5.4	Non-metric methods for pattern classification: Non-numeric data or nominal data	1
5.5	Decision trees: Classification and Regression Trees (CART) lecture 1	1
5.6	Decision trees: Classification and Regression Trees (CART) lecture 2	1

References

1. S.Theodoridis and K.Koutroumbas, "Pattern Recognition", 4th Ed., Academic Press, 2009

2. C.M.Bishop, "Pattern Recognition and Machine Learning", Springer, 2006

3. R.O.Duda, P.E.Hart and D.G.Stork, "Pattern Classification", John Wiley, 2001

4. Hastie, T., Tibshirani, R. and Friedman, J. "The Elements of Statistical Learning". Springer. 2001.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS022	ADVANCED COMPUTATIONAL LINGUISTICS	PROGRAMME ELECTIVE 4	3	0	0	3

Preamble: To familiarize with theconcepts in computational linguistics, modern grammar formalisms and NL generation. It covers TAG, Statistical parsing, NLG, Discourse processing, Lexical Functional Grammar(LFG), Morphotactics, Finite State Transducers and its use in morphology. Students will be able to understandthe basics of TAG, statistical parsing, Logical forms, LFG, well-formedness of formalisms and apply them to Natural Language Generation(NLG) and Parsing

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

CO 1	Understand the basic concepts of Statistical Parsing and Semantic Knowledge Representation. (Cognitive knowledge level: Understand)
CO 2	Make use of the concepts of Semantic Interpretation for problems (Cognitive knowledge level: Apply)
CO 3	Apply the concepts of Speech Understanding in NL problems. (Cognitive knowledge level: Apply)
CO 4	Apply the fundamentals of Lexical Functional Grammars (Cognitive knowledge level: Apply)
CO 5	Apply the concepts of Morphological Parsing in real life problems. (Cognitive knowledge level: Apply)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1			✓	✓		✓	
CO 2			✓	✓		✓	
CO 3	✓		✓	✓		✓	
CO 4	✓		✓	✓		✓	
CO 5	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	40
Analyse	20
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.
Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.
There will be two parts; Part A and Part B.
Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks
Total duration of the examination will be 150 minutes.
Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Give the details of dependency grammar.

2. How does Semantics relate to knowledge representation? What are the uses of Semantic-Based Knowledge Representation?

Course Outcome 2 (CO2)

1. Differentiate between syntactic and semantic grammar.

2. What is compositional semantics? Explain the significance of deep structure.

Course Outcome 3(CO3):

1. How does Natural Language Generation (NLG) works? What are the application areas of NLG?

2. Describe Natural Language Generation techniques.

Course Outcome 4 (CO4):

1. Describe the Lexical Functional Grammar.
2. What are the well-formedness conditions?

Course Outcome 5 (CO5):

1. What is morphotactics? Describe the importance of morphotactics?
2. How can we use Lexicon-free Finite-State Transducers for stemming?

Model Question Paper

QP CODE:

Reg No: _______________

Name: __________________ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 222CS022

Course Name:ADVANCEDCOMPUTATIONALLINGUISTICS

Max. Marks: 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1.	Differentiate Tree adjoining Grammar (TAG) from Context free Grammar (CFG)	
2.	Illustrate semantic knowledge representation with a suitable example	
3.	How effectively, dependency grammar does statistical parsing?	
4.	Illustrate how ***wh*** *movement* can be handled in questions.	
5.	What is morphological parsing? Explain with a suitable example.	**(5x5=25)**

Part B

(Answer any five questions. Each question carries 7 marks)

6.	(a)	How dependency grammars are useful in syntactic parsing? Explain.	**(4)**
	(b)	Illustrate the derivation using TAGs	**(3)**
7.	(a)	Explain case frames and case frame based parsing	**(4)**

	(b)	Explain how semantic processing contributes to language modelling	**(3)**
8.	(a)	Differentiate NLG from NLU and the challenges in them.	**(4)**
	(b)	Explain discourse processing with an example	**(3)**
9.	(a)	Differentiate inflectional morphology from dative morphology	**(4)**
	(b)	Explain Active, passive, and dative constructions.	**(3)**
10		What is a finite state transducer? How is it useful in morphological analysis?	**(7)**
11	(a)	Explain Discourse Processing and its significance	**(4)**
	(b)	Explain the techniques used in NLG	**(3)**
12	(a)	Explain how morphological parsing is done using FST	**(4)**
	(b)	Explain the significance of lexicons and morphotactics in morphological parsing	**(3)**

Syllabus

	Syllabus	
Module	**Contents**	**Hours**
I	Tree Adjoining Grammars-Dependency Grammars Statistical Parsing- Introduction to Semantic Processing, Semantic Knowledge Representation	8
II	Deep Structure and Logical Form-Compositional Semantic Interpretation. Semantic Grammars-Case Frames and Case Frame based Parsing	7
III	Natural Language Generation- Problems in NL Generation Basic Generation Techniques-Hard Problems in NLP Speech Understanding and Translation-Discourse Processing	8
IV	Lexical Functional Grammar: Active-Passive and Dative Constructions-Wh-movement in Questions-Overview of LFG-LFG Formalism: Well-formedness Conditions Handling Wh-movement in Questions-Computational Aspects	7

V	Morphology and Finite State Transducers-Inflectional Morphology-Derivational Morphology-Finite State Morphological Parsing-The Lexicon and Morphotactics Morphological Parsing with Finite State Transducers Orthographic Rules and Finite-State Transducers-Combining an FST Lexicon and Rules-Lexicon free FSTs	9

Course Plan (For 3 credit courses, the content can be for 40 hrs and for 2 credit courses, the content can be for 26 hrs. The audit course in the third semester can have content for 30 hours).

No	Topic	No. of Lectures (40)
1	**Module 1 (Semantic Processing)**	
1.1	Tree Adjoining Grammars	1
1.2	Dependency Grammars	1
1.3	Categorial Grammars	1
1.4	Statistical Parsing	1
1.5	Applications of Statistical Parsing	1
1.6	Introduction to Semantic Processing	1
1.7	Introduction to Semantic Knowledge Representation	1
1.8	Semantic Knowledge Representation	1
2	**Module 2 (Semantic Interpretation)**	
2.1	Deep Structure	1
2.2	Logical Form	1
2.3	Compositional Semantic Interpretation	1
2.4	Semantic Grammars	1
2.5	Definite Clause Grammar(DCG)	1
2.5	Case Frames	1
2.6	Introduction to Frame based Parsing	1
2.7	Frame based Parsing	1
3	**Module 3 (Speech Understanding and Translation)**	
3.1	Natural Language Generation	1
3.2	Problems in NL Generation	1
3.3	Basic Generation Techniques	1
3.4	Introduction to Hard Problems in NLP	1
3.5	Hard Problems in NLP	1
3.6	Introduction to Speech Understanding and Translation	1
3.7	Speech Understanding and Translation Explanation	1
3.8	Discourse Processing	1

4	**Module 4 (Lexical Functional Grammar)**	
4.1	Lexical Functional Grammar	1
4.2	Active-Passive and Dative Constructions	1
4.3	Wh-movement in Questions	1
4.4	Overview of LFG-LFG Formalism	1
4.5	Well-formedness Conditions	1
4.6	Handling Wh-movement in Questions	1
4.7	Computational Aspects	1
5	**Module 5 (Morphological Parsing)**	
5.1	Morphology	1
5.2	Finite State Transducers	1
5.3	Inflectional Morphology	1
5.4	Derivational Morphology	1
5.5	Finite State Morphological Parsing	1
5.6	The Lexicon and Morphotactics	1
5.7	Morphological Parsing with Finite State Transducers	1
5.8	Orthographic Rules and Finite-State Transducers	1
5.9	Combining an FST Lexicon and Rules-Lexicon free FSTs	1

Reference Books

1. Jurafsky, D. and J. H. Martin, Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall, 2008.

2. Alexander Clark, Chris Fox, and Shalom Lappin (Editors): The Handbook of Computational Linguistics and Natural Language Processing (Wiley-Blackwell Handbooks in Linguistics), 2010.

3. Akshar Bharathi, Vineet Chaitanya, and Rajeev Sangal: Natural Language Processing:APaninian Perspective. Prentice Hall of India, 1995.

4. James Allen: Natural Language Understanding. Pearson, 2002.

SEMESTER II

INTERDISCIPLINARY ELECTIVE

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS056	INTRODUCTION TO MACHINE LEARNING	INTERDISCIPLINARY ELECTIVE	3	0	0	3

Preamble: This course helps the learners to understand the concepts in Machine Learning. Students will be able to understand the basics of regression, classification and clustering. After completing this course students will be able to develop machine learning based solution for real world problems in multidisciplinary environments.

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

CO 1	Illustrate the concept, purpose, scope, steps, and applications of ML techniques. **(Knowledge level : Apply)**
CO 2	Understand the concepts of supervised, unsupervised and reinforcement learning to apply in real world problems. **(Knowledge level : Apply)**
CO 3	Illustrate the working of classifiers and clustering techniques for typical machine learning applications. **(Knowledge level : Apply)**
CO 4	Acquire skills to improve the performance of Machine Learning models using ensemble techniques. **(Knowledge level : Apply)**
CO5	Design and Implement solution for a real world problem using Machine Learning algorithms (Cognitive Knowledge Level: **Create**)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓	✓		
CO 2	✓		✓	✓	✓		
CO 3	✓		✓	✓	✓		
CO 4	✓		✓	✓	✓		
CO5	✓	✓	✓	✓	✓	✓	✓

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	80%
Analyse	20%
Evaluate	
Create	

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Continuous Internal Evaluation : 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/Quiz : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted.

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

Total : 60 marks

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the precision and recall for the data.

2. Distinguish between supervised learning and Reinforcement learning. Illustrate with an example.

3. Discuss any four examples of machine learning applications.

Course Outcome 2 (CO2)

1. State the mathematical formulation of the SVM problem. Give an outline of the method for solving the problem.

2. Show the final result of hierarchical clustering with complete link by drawing a dendrogram.

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.70	0.45	0	
F	0.34	0.61	0.93	0.20	0.67	0

Course Outcome 3(CO3):

1. Identify the first splitting attribute for the decision tree by using the ID3 algorithm with the following dataset.

Major	Experience	Tie	Hired?
CS	programming	pretty	NO
CS	programming	pretty	NO
CS	management	pretty	YES
CS	management	ugly	YES
business	programming	pretty	YES
business	programming	ugly	YES
business	management	pretty	NO
business	management	pretty	NO

2. Consider the training data in the following table where Play is a class attribute. In the table, the Humidity attribute has values "L" (for low) or "H" (for high), Sunny has values "Y" (for yes) or "N" (for no), Wind has values "S" (for strong) or "W" (for weak), and Play has values "Yes" or "No".

Humidity	Sunny	Wind	Play
L	N	S	No
H	N	W	Yes
H	Y	S	Yes
H	N	W	Yes
L	Y	S	No

What is the class label for the following day (Humidity=L, Sunny=N, Wind=W), according to naïve Bayesian classification?

3. Explain DBSCAN algorithm for density based clustering. List out its advantages compared to K-means.

4. Explain how Support Vector Machine can be used for classification of linearly separable data.

5. Define Hidden Markov Model. What is meant by the evaluation problem and how is this solved?

6. Use K Means clustering to cluster the following data into two groups. Assume cluster centroid are m1=2 and m2=4. The distance function used is Euclidean distance. { 2, 4, 10, 12, 3, 20, 30, 11, 25 }

Course Outcome 4 (CO4):

1. Explain how the *Random Forests* give output for *Classification*, and *Regression* problems?
2. Is *Random Forest* an *Ensemble Algorithm*

3. Why is the training efficiency of Random Forest better than Bagging?

Model Question Paper

Reg No: _______________

Name: _________________ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: 222ECS056

Course Name: INTRODUCTION TO MACHINE LEARNING

Max. Marks: 60 **Duration: 2.5 Hours**

PART A

Answer All Questions. Each Question Carries 5 Marks

1.	Bias-Variance trade-off is a design consideration while training the machine learning model. Justify.	
2.	Derive the expression for sigmoid function associated with Logistic Regression.	
3.	How Optimal Marginal Hyperplane contributes to the accuracy of predictions using SVM. Justify how Kernel functions are used in Linear Inseparable problems	
4.	Discuss how good DBSCAN is in clustering data points available in dense Euclidean space.	
5.	Suggest an ensemble method that generates one classifier per round	**(5x5=25)**

Part B

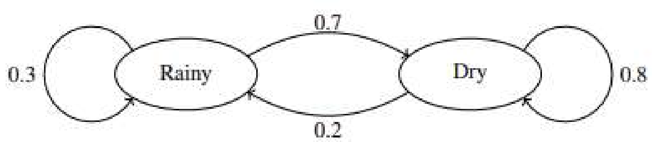
(Answer any five questions. Each question carries 7 marks)

6.		Explain various Cost Functions associated with Regression & Classification.	**(7)**
7.		Weight updation contributes to the performance of a Neural Network model. Justify the statement using the Back propagation algorithm.	**(7)**
8.		Compute the Principal Components for the 2D data: X=(x1,x2)={(1,2),(3,3),(3,5),(5,4),(5,6),(6,5),(8,7),(9,8)}	**(7)**
9.		Using Naïve Bayes algorithm, predict whether a Red color car which is imported as a Sports category will be stolen or not.	**(7)**

Color	Type	Origin	Stolen?
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

10		Construct Dendrograms based on Complete Linkage and Average Linkage.(7)	**(7)**
11		Perform k-means algorithm on the data given in qn.8. (Given no. of clusters =2, iterations=2).	**(7)**
12	(a)	If P(Rain) = 0.4 and P(Dry) = 0.6 compute the probability for the sequence "Rain, Rain, Dry, Dry".	**(3)**
	(b)	Elucidate the three basic problems of HMM	**(4)**

ID	1	2	3	4	5
1	0	3	18	10	25
2	3	0	21	13	28
3	18	21	0	8	7
4	10	13	8	0	15
5	25	28	7	15	0



	Syllabus	
Module	**Contents**	**Hours**
I	Overview of machine learning: supervised, semi-supervised, unsupervised learning, reinforcement learning. Types of ML problems: Classification, Clustering and Regression, Cost functions: Definition and Types, Data PreProcessing, Bias-Variance trade off, Cross validation techniques, Classifier performance measures, ROC Curves	6
II	Introduction to neural network : Linear Regression, Least square Gradients, Logistic Regression, Sigmoid function & differentiation, Logistic Regression – Regularization, Neural Networks – Concept of perceptron and Artificial neuron, Weight initialization techniques, Feed Forward Neural Network, Back Propagation algorithm	8
III	Classification Methods : Support Vector Machine, Optimal Separating hyper plane, Kernel trick, Kernel functions, Gaussian class conditional distribution, Bayes Rule, Naïve Bayes Model, Decision Tree – ID3, Maximum Likelihood estimation techniques	8

IV	Clustering Methods: K-means clustering , Hierarchical clustering techniques, Density Based clustering, Feature Selection techniques: Entropy, Correlation Coefficient, Chi-square Test, Forward & Backward Selection, Dimensionality Reduction: PCA, LDA, t-SNE	7
V	Basics of graphical models - Bayesian networks, Hidden Markov model, Ensemble methods – Boosting, Bagging, Random forest, XGBoost (Case study)	6

Lesson Plan

1	**Introduction to machine learning (Hours: 6)**	
1.1	Overview of machine learning: supervised, semi-supervised, unsupervised learning, reinforcement learning	1
1.2	Types of ML problems: Classification, Clustering and Regression	1
1.3	Cost functions: Definition and Types	1
1.4	Data PreProcessing, Bias-Variance trade off	1
1.5	Cross validation techniques	1
1.6	Classifier performance measures, ROC Curves	1
2	**Introduction to neural network (Hours: 8)**	
2.1	Linear Regression, Least square Gradients	1
2.2	Logistic Regression	1
2.3	Sigmoid function & differentiation	1
2.4	Logistic Regression - Regularization	
2.5	Neural Networks – Concept of perceptron and Artificial neuron	1
2.6	Weight initialization techniques	1
2.7	Feed Forward Neural Network	1
2.8	Back Propagation algorithm	1
3	**Classification Methods (Hours: 8)**	
3.1	Support Vector Machine	1
3.2	Optimal Separating hyper plane	1
3.3	Kernel trick, Kernel functions	1
3.4	Gaussian class conditional distribution	1
3.5	Bayes Rule	1
3.6	Naïve Bayes Model	1
3.7	Decision Tree – ID3,	1
3.8	Maximum Likelihood estimation techniques	1
4	**Clustering Methods (Hours: 7)**	
4.1	K-means clustering	1
4.2	Hierarchical clustering techniques	1
4.3	BIRCH	1
4.4	Density Based clustering	1
4.5	Feature Selection techniques: Entropy, Correlation Coefficient, Chi-square Test	1
4.6	Forward & Backward Selection	1
4.5	Dimensionality Reduction: PCA	1
4.6	LDA	1

4.7	t-SNE	1
5	**Basics of graphical models (Hours: 6)**	
5.1	Basics of graphical models - Bayesian networks	1
5.2	Hidden Markov model	1
5.3	Ensemble methods - Boosting	1
5.4	Bagging	1
5.5	Random forest	1
5.6	XGBoost (Case study)	1

Reference Books

1. Ethem Alpaydın, "Introduction to Machine Learning (Adaptive Computation and Machine Learning)", MIT Press, 2004.

2. Kevin Murphy, Machine Learning: A Probabilistic Perspective (MLAPP), MIT Press, 2012

3. Han, Jiawei, and Micheline Kamber. Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers

4. Christopher M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS057	DATA STRUCTURES	INTERDISCIPLINARY ELECTIVE	3	0	0	3

Preamble: The purpose of the syllabus is to create awareness about Data Structures and their applications. After the completion of the course, the learners should be able to either use existing data structures or design their own data structures to solve real world problems.

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

CO1	Design algorithms for a task and calculate the time complexity of that algorithm (**Cognitive Knowledge Level: Apply**)
CO2	Use arrays and linked lists for problem solving (Cognitive Knowledge Level: **Apply**)
CO3	Represent data using trees, graphs and manipulate them to solve computational problems. (**Cognitive Knowledge Level:Apply**)
CO4	Make use of appropriate sorting algorithms to order data based on the situation. (**Cognitive Knowledge Level: Apply**)
CO5	Design and Implement appropriate Data Structures for solving a real world problem (Cognitive Knowledge Level: **Create**)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	✓		✓	✓		✓	
CO2	✓		✓	✓		✓	
CO3	✓		✓	✓		✓	
CO4	✓		✓	✓		✓	
CO5	✓	✓	✓	✓		✓	✓

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	70%
Analyse	30%
Evaluate	
Create	

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:
The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Sample Course Level Assessment Questions

Course Outcome1(CO1):
1. Write an algorithm for matrix multiplication and calculate its time complexity.
2. Define Big-O notation. Derive the Big – O notation for $5n^3+2n^2+3^n$.
3. Check whether the following is true or not.
   2n+1 is O(2n). Give reason.

Course Outcome 2(CO2):
1. How a linked list can be used to represent the polynomial

$5x^4y^6+24x^3y^4-17x^2y^3+15xy^2+45$.

Write a procedure to add two Bivariate polynomials represented using linked lists.

2. Write an algorithm/pseudocode to convert a given infix expression to postfix expression. Trace the steps involved in converting the given infix expression ((A +B)^C)-((D*C)/F) to postfix expression.

3. Let L1 be a singly linked list in memory. Write an algorithm that
i) Finds the number of non zero elements in L1
ii) Adds a given value K to each element in L1

Course Outcome 3(CO3):

1. Create a Binary Tree with the following sequence 14, 15, 4, 18, 9, 16, 20, 17, 3, 7, 5, 2 and perform inorder, preorder and postorder traversals on the above tree and print the output of the traversals.

2. In a complete binary tree of depth d (complete including last level), give an expression to find the number of leaf nodes.

Course Outcome 4(CO4):

1. Write an algorithm/pseudocode to sort elements using Heap sort technique. Illustrate the working of Heap sort algorithm on the following input : 35,15,0,1,60

2. With the help of an algorithm/pseudocode and suitable example, explain how you would perform binary search on an array of n elements. Find the time complexity of binary search algorithm.

3. Suppose an array contains elements {10, 13, 21, 32, 35, 44, 55}. Give the steps to find an element "35" using i) linear search ii) binary search

Course Outcome 5(CO5):

Design a reservation system for railways that includes a waiting list. If the reservation is full, display "reservation full" and put them in the waiting list and give a waiting list number. If a passenger wishes to cancel his ticket, he may do it any time. Then the passenger at the front of the waiting list is allotted a berth automatically.

Model Question Paper

Reg No: _______________			
Name: _________________			**PAGES : 4**
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY			
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR			
Course Code: 222CS057			
Course Name: Data Structures			
Max. Marks: 60			**Duration: 2.5 Hours**
PART A			
Answer All Questions. Each Question Carries 5 Marks			
1.		Write an algorithm to add a new element in a particular position of an array.	
2.		Compare Circular Queue with a Normal Queue	
3.		How can a doubly linked list be used to find palindromes?	
4.		Write an iterative algorithm for in-order traversal of a Binary Tree	
5.		Trace the working of Quick sort on the following input 38,8,0,28,45,-12,89,66,42.	**(5x5=25)**
Part B			
(Answer any five questions. Each question carries 7 marks)			
6.	(a)	How is the performance of an algorithm evaluated?	**(4)**
	(b)	In the functions O(nlogn) and O(logn), which one is better in terms of computational complexity and why?	**(3)**
7.	(a)	Write algorithms to insert and delete elements from a double ended queue. Illustrate with examples	**(7)**
8.	(a)	Write an algorithm to multiply two polynomials represented using linked list	**(7)**
9.	(a)	List the properties of Binary Search Tree. Write an algorithm to search an element in a Binary Search Tree	**(7)**
10		Give algorithms for DFS and BFS of a graph and explain with examples.	**(7)**
11	(a)	Write and illustrate algorithms for Merge Sort and Quick Sort.	**(7)**
12	(a)	How is memory de-allocation done in memory management	**(7)**

	Syllabus	
Module	**Contents**	**Hours**
I	**Basic Concepts of Data Structures** System Life Cycle, Algorithms, Performance Analysis, Space complexity, Time Complexity, Asymptotic Notation, Complexity Calculation of Simple Algorithms	5
II	**Arrays and Searching** Polynomial representation using Arrays, Sparse matrix, Stacks, Queues, Circular Queues, Priority Queues, Double Ended Queues, Evaluation of Expressions, Linear Search and Binary Search	9
III	**Linked List and Memory Management** Self Referential Structures, Dynamic Memory Allocation, Singly Linked List-Operations on Linked List, Doubly Linked List, Circular Linked List, Stacks and Queues using Linked List, Polynomial representation using Linked List, Memory allocation and de-allocation,	8
IV	Trees, Binary Trees-Tree Operations, Binary Tree Representation, Tree Traversals, Graphs- Representation of Graphs, Depth First Search and Breadth First Search on Graphs, Applications of Graphs	7
V	Searching Techniques – Linear search , Binary search Sorting Techniques – Selection Sort, Insertion Sort, Quick Sort, Merge Sort and Heap Sort	5

Course Plan

Module 1 :Basic Concepts of Data Structures		(5 hours)
1.1	System Life Cycle,	1 hour
1.2	Algorithms	1 hour
1.3	Performance Analysis, Space Complexity, Time Complexity,	1 hours
1.4	Asymptotic Notation	1hour
1.5	Complexity Calculation of Simple Algorithms	1hour
Module 2 :Arrays and Searching		(9 hours)
2.1	Polynomial representation using Arrays	1 hour
2.2	Sparse matrix	1 hours

2.3	Stacks	1 hour
2.4	Queues, Circular Queues	1 hour
2.5	Priority Queues,	1 hour
2.6	Double Ended Queue	1 hours
2.7	Evaluation of Expressions	1 hour
2.8	Linear Search	1 hour
2.9	Binary Search	1 hour
Module 3 : Linked List and Memory Management		(8 hours)
3.1	Self Referential Structures	1 hour
3.2	Dynamic Memory Allocation	1 hour
3.3	Single Linked List-Operations on Linked List,	1 hour
3.4	Double Linked List	1 hour
3.5	Circular Linked List	1 hour
3.6	Stacks and Queues using Linked List	1 hour
3.7	Polynomial representation using Linked List	1 hour
3.8	Memory de-allocation	1 hour
Module 4 :Trees and Graphs		(7 hours)
4.1	Trees, Binary Trees	1hour
4.2	Tree Operations, Binary Tree Representation,	1hour
4.3	Tree Traversals	1hour
4.4	Graphs	1hour
4.5	Representation of Graphs	1hour
4.6	Depth First Search and Breadth First Search on Graphs	1hour
4.7	Applications of Graphs	1hour
Module 5 : Sorting and Hashing		(5 hours)
5.1	Sorting Techniques – Selection Sort	1hour
5.2	Insertion Sort	1hour
5.3	Quick Sort	1hour
5.4	Merge Sort	1hour
5.5	Heap Sort	1hour

Text Book

1. Ellis Horowitz, Sartaj Sahni and Susan Anderson-Freed, Universities Press, Fundamentals of Data Structures in C

Reference Books

1. Samanta D., Classic Data Structures, Prentice Hall India, 2/e, 2009.
2. Richard F. Gilberg, Behrouz A. Forouzan, Data Structures: A Pseudocode Approach with C, 2/e, Cengage Learning, 2005
3. Aho A. V., J. E. Hopcroft and J. D. Ullman, Data Structures and Algorithms, Pearson Publication,1983.
4. Tremblay J. P. and P. G. Sorenson, Introduction to Data Structures with Applications, Tata McGraw Hill, 1995.
5. Peter Brass, Advanced Data Structures, Cambridge University Press, 2008
6. Lipschuts S., Theory and Problems of Data Structures, Schaum's Series, 1986.
7. Wirth N., Algorithms + Data Structures = Programs, Prentice Hall, 2004.
8. Hugges J. K. and J. I. Michtm, A Structured Approach to Programming, PHI, 1987.
9. Martin Barrett, Clifford Wagner, And Unix: Tools For Software Design, John Wiley, 2008 reprint

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222ECS058	SOFTWARE PROJECT MANAGEMENT	INTERDISCIPLINARY ELECTIVE	3	0	0	3

Preamble: This course provides fundamental knowledge in the Software Development Process. It covers Software Development, Quality Assurance, Project Management concepts and technology trends. This course enables the learners to apply state of the art industry practices in Software development.

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

CO1	Demonstrate Traditional and Agile Software Development approaches **(Cognitive Knowledge Level: Apply)**
CO2	Prepare Software Requirement Specification and Software Design for a given problem. **(Cognitive Knowledge Level**: **Apply**)
CO3	Justify the significance of design patterns and licensing terms in software development, prepare testing, maintenance and DevOps strategies for a project. **(Cognitive Knowledge Level: Apply**)
CO4	Make use of software project management concepts while planning, estimation, scheduling, tracking and change management of a project, with a traditional/agile framework**. (Cognitive Knowledge Level: Apply**)
CO5	Utilize SQA practices, Process Improvement techniques and Technology advancements in cloud based software models and containers & microservices. **(Cognitive Knowledge Level: Apply**)

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: An ability to apply stream knowledge to design or develop solutions for real world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources and state-of-the-art tool to model, analyse and solve practical engineering problems.

PO6: An ability to engage in life-long learning for the design and development related to the stream related problems taking into consideration sustainability, societal, ethical and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓		✓	
CO 2	✓		✓	✓		✓	
CO 3	✓		✓	✓		✓	
CO 4	✓		✓	✓		✓	
CO 5	✓		✓	✓		✓	

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%
Analyse	30%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design based questions (for both internal and end semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. What are the advantages of an incremental development model over a waterfall model?
2. Illustrate how the process differs in agile software development and traditional software development with a socially relevant case study. (Assignment question)

Course Outcome 2 (CO2):

1. How to prepare a software requirement specification?
2. Differentiate between Architectural design and Component level design.
3. How does agile approaches help software developers to capture and define the user requirements effectively?
4. What is the relevance of the SRS specification in software development?
5. Prepare a use case diagram for a library management system.

Course Outcome 3 (CO3):

1. Differentiate between the different types of software testing strategies.
2. Justify the need for DevOps practices?
3. How do design patterns help software architects communicate the design of a complex system effectively?
4. What are the proactive approaches one can take to optimise efforts in the testing phase?

Course Outcome 4 (CO4):

1. Illustrate the activities involved in software project management for a socially relevant problem?
2. How do SCRUM, Kanban and Lean methodologies help software project management?
3. Is rolling level planning in software project management beneficial? Justify your answer.
4. How would you assess the risks in your software development project? Explain how you can manage identified risks?

Course Outcome 5 (CO5):

1. Justify the importance of Software Process improvement?
2. Explain the benefits of cloud based software development, containers and microservices.
3. Give the role of retrospectives in improving the software development process.
4. Illustrate the use of project history data as a prediction tool to plan future socially relevant projects.

Model Question Paper

Reg No: _______________			
Name: _________________			**PAGES : 4**
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY			
SECOND SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR			
Course Code: 222CS058			
Course Name: SOFTWARE PROJECT MANAGEMENT			
Max. Marks: 60			**Duration: 2.5 Hours**
PART A			
Answer All Questions. Each Question Carries 5 Marks			
1.	Explain Agile ceremonies and Agile manifesto.		
2.	Compare Software Architecture design and Component level design		
3.	Describe the formal and informal review techniques		
4.	Explain plan driven development and project scheduling.		
5.	Illustrate SPI process with an example.		**(5x5=25)**
Part B **(Answer any five questions. Each question carries 7 marks)**			
6.		Illustrate software process activities with an example.	**(7)**
7.		What are functional and nonfunctional requirements? Imagine that you are developing a library management software for your college, list eight functional requirements and four nonfunctional requirements.	**(7)**

8.		Explain Continuous Integration, Delivery, and Deployment CI/CD/CD)	**(7)**
9.		What is a critical path and demonstrate its significance in a project schedule with the help of a sample project schedule.	**(7)**
10.		What is algorithmic cost modeling? What problems does it suffer from when compared with other approaches to cost estimation?	**(7)**
11.		Explain elements of Software Quality Assurance and SQA Tasks.	**(7)**
12.		Compare CMMI and ISO 9001:2000.	**(7)**

Syllabus

Module 1 : Introduction to Software Engineering (7 hours)

Introduction to Software Engineering - Professional software development, Software engineering ethics. Software process models - The waterfall model, Incremental development. Process activities - Software specification, Software design and implementation, Software validation, Software evolution. Coping with change - Prototyping, Incremental delivery, Boehm's Spiral Model. Agile software development - Agile methods, agile manifesto - values and principles. Agile development techniques, Agile Project Management. Case studies : An insulin pump control system. Mentcare - a patient information system for mental health care.

Module 2 : Requirement Analysis and Design (8 hours)

Functional and non-functional requirements, Requirements engineering processes. Requirements elicitation, Requirements validation, Requirements change, Traceability Matrix. Developing use cases, Software Requirements Specification Template, Personas, Scenarios, User stories, Feature identification. Design concepts - Design within the context of software engineering, Design Process, Design concepts, Design Model. Architectural Design - Software Architecture, Architectural Styles, Architectural considerations, Architectural Design Component level design
- What is a component?, Designing Class-Based Components, Conducting Component level design, Component level design for web-apps. Template of a Design Document as per “IEEE Std 1016-2009 IEEE Standard for Information Technology Systems Design Software Design Descriptions”. Case study: The Ariane 5 launcher failure.

Module 3 : Implementation and Testing (9 hours)

Object-oriented design using the UML, Design patterns, Implementation issues, Open-source development - Open-source licensing - GPL, LGPL, BSD. Review Techniques - Cost impact of Software Defects, Code review and statistical analysis. Informal Review, Formal Technical Reviews, Post-mortem evaluations. Software testing strategies - Unit Testing, Integration Testing, Validation testing, System testing, Debugging, White box testing, Path testing, Control Structure testing, Black box testing, Testing Documentation and Help facilities. Test automation, Test-driven development, Security testing. Overview of DevOps and Code Management - Code management, DevOps automation, Continuous Integration, Delivery, and Deployment (CI/CD/CD). Software Evolution - Evolution processes, Software maintenance.

Module 4 : Software Project Management (6 hours)

Software Project Management - Risk management, Managing people, Teamwork. Project Planning, Software pricing, Plan-driven development, Project scheduling, Agile planning. Estimation techniques, COCOMO cost modeling. Configuration management, Version management, System building, Change management, Release management, Agile software management - SCRUM framework. Kanban methodology and lean approaches.

Module 5 : Software Quality, Process Improvement and Technology trends (6 hours)

Software Quality, Software Quality Dilemma, Achieving Software Quality Elements of Software Quality Assurance, SQA Tasks , Software measurement and metrics. Software Process Improvement(SPI), SPI Process CMMI process improvement framework, ISO 9001:2000 for Software. Cloud-based Software - Virtualisation and containers, Everything as a service(IaaS, PaaS), Software as a service. Microservices Architecture - Microservices, Microservices architecture, Microservice deployment.

Text Books

1. Book 1 - Ian Sommerville, Software Engineering, Pearson Education, Tenth edition, 2015.
2. Book 2 - Roger S. Pressman, Software Engineering : A practitioner’s approach, McGraw Hill publication, Eighth edition, 2014
3. Book 3 - Ian Sommerville, Engineering Software Products: An Introduction to Modern Software Engineering, Pearson Education, First Edition, 2020.

References

2. IEEE Std 830-1998 - IEEE Recommended Practice for Software Requirements SpeciÞcations
3. IEEE Std 1016-2009 IEEE Standard for Information Technology—Systems Design— Software Design Descriptions
4. David J. Anderson, Kanban, Blue Hole Press 2010

5. David J. Anderson, Agile Management for Software Engineering, Pearson, 2003
6. Walker Royce, Software Project Management : A unified framework, Pearson Education, 1998
7. Steve. Denning, The age of agile, how smart companies are transforming the way work gets done. New York, Amacom, 2018.
8. Satya Nadella, Hit Refresh: The Quest to Rediscover Microsoft's Soul and Imagine a Better Future for Everyone, Harper Business, 2017
9. Henrico Dolfing, Project Failure Case Studies: Lessons learned from other people's mistakes, Kindle edition
10. Mary Poppendieck, Implementing Lean Software Development: From Concept to Cash, Addison-Wesley Signature Series, 2006
11. StarUML documentation - https://docs.staruml.io/
12. OpenProject documentation - https://docs.openproject.org/
13. BugZilla documentation - https://www.bugzilla.org/docs/
14. GitHub documentation - https://guides.github.com/
15. Jira documentation - https://www.atlassian.com/software/jira

Teaching Plan

No	Contents	No of Lecture Hrs
Module 1 : Introduction to Software Engineering (7 hours)		
1.1	Introduction to Software Engineering.[Book 1, Chapter 1]	1 hour
1.2	Software process models [Book 1 - Chapter 2]	1 hour
1.3	Process activities [Book 1 - Chapter 2]	1 hour
1.4	Coping with change [Book 1 - Chapter 2, Book 2 - Chapter 4]	1 hour
1.5	Case studies : An insulin pump control system. Mentcare - a patient information system for mental health care. [Book 1 - Chapter 1]	1 hour
1.6	Agile software development [Book 1 - Chapter 3]	1 hour
1.7	Agile development techniques, Agile Project Management.[Book 1 - Chapter 3]	1 hour
Module 2 : Requirement Analysis and Design (8 hours)		
2.1	Functional and non-functional requirements, Requirements engineering processes [Book 1 - Chapter 4]	1 hour

2.2	Requirements elicitation, Requirements validation, Requirements change, Traceability Matrix [Book 1 - Chapter 4]	1 hour
2.3	Developing use cases, Software Requirements Specification Template [Book 2 - Chapter 8]	1 hour
2.4	Personas, Scenarios, User stories, Feature identification [Book 3 - Chapter 3]	1 hour
2.5	Design concepts [Book 2 - Chapter 12]	1 hour
2.6	Architectural Design [Book 2 - Chapter 13]	1 hour
2.7	Component level design [Book 2 - Chapter 14]	1 hour
2.8	Design Document Template. Case study: The Ariane 5 launcher failure. [Ref - 2, Book 2 - Chapter 16]	1 hour
Module 3 : Implementation and Testing (9 hours)		
3.1	Object-oriented design using the UML, Design patterns [Book 1 - Chapter 7]	1 hour
3.2	Implementation issues, Open-source development - Open-source licensing - GPL, LGPL, BSD [Book 1 - Chapter 7]	1 hour
3.3	Review Techniques - Cost impact of Software Defects, Code review and statistical analysis. [Book 2 - Chapter 20]	1 hour
34	Informal Review, Formal Technical Reviews, Post-mortem evaluations. [Book 2 - Chapter 20]	1 hour
3.5	Software testing strategies - Unit Testing, Integration Testing, Validation testing, System testing and Debugging (basic concepts only). [Book 2 - Chapter 22]	1 hour
3.6	White box testing, Path testing, Control Structure testing, Black box testing. Test documentation [Book 2 - Chapter 23]	1 hour
3.7	Test automation, Test-driven development, Security testing. [Book 3 - Chapter 9]	1 hour
3.8	DevOps and Code Management - Code management, DevOps automation, CI/CD/CD. [Book 3 - Chapter 10]	1 hour
3.9	Software Evolution - Evolution processes, Software maintenance. [Book 1 - Chapter 9]	1 hour
Module 4 : Software Project Management (6 hours)		

4.1	Software Project Management - Risk management, Managing people, Teamwork [Book 1 - Chapter 22]	1 hour
4.2	Project Planning - Software pricing, Plan-driven development, Project scheduling, Agile planning [Book 1 - Chapter 23]	1 hour
4.3	Estimation techniques [Book 1 - Chapter 23]	1 hour
4.4	Configuration management [Book 1 - Chapter 25]	1 hour
4.5	Agile software management - SCRUM framework [Book 2 - Chapter 5]	1 hour
4.6	Kanban methodology and lean approaches.[Ref 9 - Chapter 2]	1 hour
Module 5 : Software Quality, Process Improvement and Technology trends (6 hours)		
5.1	Software Quality, Software Quality Dilemma, Achieving Software Quality. [Book 2 - Chapter 19]	1 hour
5.2	Elements of Software Quality Assurance, SQA Tasks , Software measurement and metrics. [Book 3 - Chapter 21]	1 hour
5.3	Software Process Improvement (SPI), SPI Process [Book 2 - Chapter 37]	1 hour
5.4	CMMI process improvement framework, ISO 9001:2000 for Software. [Book 2 - Chapter 37]	1 hour
5.5	Cloud-based Software - Virtualisation and containers, IaaS, PaaS, SaaS.[Book 3 - Chapter 5]	1 hour
5.6	Microservices Architecture - Microservices, Microservices architecture, Microservice deployment [Book 3 - Chapter 6]	1 hour

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
222EEX100	INTRODUCTION TO BIG DATA ANALYTICS AND DATA ENGINEERING	INDUSTRY ELECTIVE	3	0	0	3

Preamble: This course is intended to articulate the main concepts of Big data, Hadoop, and Elastic Search. This course helps students learn and apply important tools for managing massive data, namely Hadoop, MapReduce, Elasticsearch, and the Elastic Stack. This course helps students to utilize elastic search and elastic stack to design the best systems for analyzing Big data and be the most valuable engineer they can be.

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

CO1	Design distributed systems that manage "big data" using Hadoop and related data engineering technologies. (**Cognitive knowledge level: Apply)**
CO2	Utilize HDFS and MapReduce for storing and analyzing data at scale. (**Cognitive knowledge level: Apply)**
CO3	Utilize Logstash and the "ELK stack" to import streaming log data into Elasticsearch (**Cognitive knowledge level: Apply)**
CO4	Analyze and visualize data in Elasticsearch using Kibana (**Cognitive knowledge level: Apply)**
CO5	Design, Develop, implement, and present a complete procedure for collecting, aggregating, monitoring, visualizing, and analyzing big data **(Cognitive Knowledge Level: create)**

Program Outcomes (PO)

Outcomes are the attributes that are to be demonstrated by a graduate after completing the course.

PO1: An ability to independently carry out research/investigation and development work in engineering and allied streams

PO2: An ability to communicate effectively, and write and present technical reports on complex engineering activities by interacting with the engineering fraternity and with society at large.

PO3: An ability to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor's program

PO4: An ability to apply stream knowledge to design or develop solutions for real-world problems by following the standards

PO5: An ability to identify, select and apply appropriate techniques, resources, and state-of-the-art tools to model, analyze and solve practical engineering problems.

PO6: An ability to engage in lifelong learning for the design and development related to stream- related problems taking into consideration sustainability, societal, ethical, and environmental aspects

PO7: An ability to develop cognitive load management skills related to project management and finance, which focus on Entrepreneurship and Industry relevance.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	✓		✓	✓	✓	✓	
CO 2	✓		✓	✓	✓	✓	
CO 3	✓			✓		✓	
CO 4	✓		✓	✓	✓	✓	
CO 5	✓	✓		✓	✓	✓	✓

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	60-80%
Analyse	20-40%
Evaluate	Assignments/projects
Create	Assignments/projects

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

The evaluation shall only be based on application, analysis, or design-based questions (for internal and end-semester examinations).

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer-reviewed original publications (minimum ten publications shall be referred) : 15 marks
ii. Course-based task / Seminar/ Data collection and interpretation : 15 marks
iii. Test paper (1 number) : 10 marks

The test paper shall include a minimum of 80% of the syllabus.

Course-based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation, and understanding of the students.

End Semester Examination Pattern:

The end-semester examination will be conducted by the respective College.

There will be two parts: Part A and Part B.
Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem-solving and quantitative evaluation), with a minimum one question from each module of which student should answer any five. Each question can carry 7 marks.
The total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.
For example, if the average end-semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Explain pre-built summary tables and on-demand summary tables with examples. For a COVID-19 dashboard, which of them will be preferred?
2. a. Elaborate on the deficiencies of conventional databases when used with Big Data.
   b. Justify: businesses are using Big Data for competitive advantage.
3. a. Explain Big Data, and where does it come from? How does it work?
   b. Explain the 5vs of Big Data

Course Outcome 2 (CO2):

1. Explain the working of HDFS and MapReduce
2. Utilize MapReduce to Find the Top k Rated Movies
3. Travel and Tourism Analysis Data of Hotels using Big Data Hadoop.

Course Outcome 3(CO3):

Assume that you have already created your Elastic search Service cluster. Once you create a cluster, you will be provided with both a Cloud ID and password for the elastic superuser account.

Use Logstash to ingest data into an elastic search using Logstash

1. Write an overview of Logstash
2. Explain the Logstash pipeline
3. Present a sample Logstash pipeline that
   a. Reads the Elastic Blogs RSS feed
   b. Performs some light pre-processing of the data by copying/renaming fields and removing special characters and HTML tags
   c. Ingests the documents into Elasticsearch

Course Outcome 4 (CO4):

1. Apply Elastic Stack for Log Management, Analysis & Analytics.
2. while troubleshooting performance issues, it is observed that a couple of machines are slowed down and reaching high-CPU utilization. This might mean that it lacks resources because of high load, but very often, there is a bug in the code, an exception, or an error flow that over-utilizes resources. Apply ELK stack to view and monitor performance metrics combined with all the events generated by the apps, operating systems, and network devices.

Course Outcome 5 (CO5):

1. Build pipelines that process and manipulate thousands of events using Logstash and Filebeat
2. Build a fully functional pipeline that handles Apache web server logs
3. Send data to Logstash from numerous sources and to several destinations

Model Question Paper			
QP CODE:			
Reg No: ______________			
Name:________________			**PAGES: 4**
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY			
FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR			
Course Code: 222EEX100			
Course Name: INTRODUCTION TO BIG DATA ANALYTICS AND DATA ENGINEERING			
Max. Marks: 60			**Duration: 2.5 Hours**
PART A			
Answer All Questions. Each Question Carries 5 Marks			
1.		Explain deploying a Big Data Model. Mention the key steps involved	**5**
2.		Contrast HDFS and Traditional NFS	**5**
3.		Distinguish between the physical plan and the logical plan in Pig script.	**5**
4.		Given the Shakespeare plays, find the longest plays, and visualize them based on how many lines are in each play. Suppose that, in the Shakespeare Index, every line of the play corresponds to a document. Create a visualization for a vertical bar chart that aggregates the document count by the play name in descending order.	**5**
5.		Explain the kibana canvas	**5** **(5x5=25)**
Part B **(Answer any five questions. Each question carries 7 marks)**			
6.	(a)	Explain Big Data Lifecycle.	**(10)**
7.	(a)	Explain the use of different types of sandboxes	**(5)**
	(b)	For a university management system, prepare a rough design for each type of sandbox.	**(5)**
8.	(a)	Illustrate map reduce job execution flow.	**(10)**
9.	(a)	Assuming you are given a dataset containing grades of all students under APJAKTU since 2015, how do you write a Pig Latin script to generate a report of the pass percentage of each college for each year? Explain the script in detail.	**(10)**
10	(a)	Create a search index of the complete works of William Shakespeare using an Elastic search	**(10)**
11	(a)	Analyze the Log data with the Elastic Stack	**(7)**
	(b)	Illustrate ELK stack Architecture	**(3)**
12	(a)	Build a system to collect Logs with Filebeat	**(10)**

Syllabus

Module No	Contents	Hours
I	**MODULE 1 Introduction to big data:** Introduction to Big Data Platform, Challenges of Conventional Systems - Intelligent data, analysis, Nature of Data - Analytic Processes and Tools - Analysis vs. Reporting.	8
II	**MODULE 2: Components of Hadoop** - Analysing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS Basics **Developing a Map Reduce Application**-Anatomy of a Map Reduce Job, Scheduling-Shuffle, and Sort-Task execution.	**8**
III	**MODULE 3:Introduction to HBase,** Filesystems for HBase, Client API - The Basics, HBase clients – REST, Shell Commands, Map Reduce Integration.**Introduction to Pig,** Grunt, pig data model, Pig Latin, Advanced pig Latin, developing and testing Pig Latin scripts, Map Reduce Integration	**8**
IV	**MODULE 4:Introduction to ELK**, general architecture of ELK, Elastic search, Installation and configuration of elastic search, Cluster setup in elastic search, Indexing, Mapping, Searching, Populate elastic search, Elasticsearch API Conventions, Elasticsearch Document APIs, Search APIs in Elasticsearch, DSL queries, Analysis in Elasticsearch, Check the performance of Elasticsearch, API available in elastic search, Create filter, get data and put data, check the health of elastic search.	**8**
V	**MODULE 5: Logstash & Kibana**, **Logstash**: Installation, and Configuration of Logstash Integrating logstash with Kibana, Collection, and Parsing of Logs, log shipping, shipping tools, filebeat, configure different filebeat, integrate filebeat to Logstash.**Kibana:** Installation of Kibana, Interact filebeat, Logstash elastic search, and Kibana, create an alert in Kibana, docker role in ELK, migrate ELK to Docker	**8**

Course Plan

S.NO	TOPIC	NO. OF LECTURES
MODULE1 -Introduction to Big Data -9 HRS		
1.1	Introduction to Big Data, Characteristics of bigdata, Bigdata Lifecycle	1
1.2	Challenges of Conventional System	1
1.3	Use cases of BDA, Big data challenges	1
1.4	Big data platform, Bigdata tools, Components of Bigdata platform	1
1.5	Types of Bigdata Analytics, Nature of data	1
1.6	Intelligent data Analysis, IDA Applications	1
1.7	Types of data in data science	1
1.8	Analytic Processes and Tools	1
MODULE 2-Hadoop & MapReduce- 8 HRS		
2.1	Introduction to Hadoop, Features of Hadoop, Components of Hadoop ecosystem	1
2.2	HDFS Architecture	1
2.3	Design of HDFS-Java interfaces to HDFS Basics	1
2.4	Introduction to MapReduce, MapReduce data flow, Mapreduce Wordcount example	1
2.5	Analyzing the Data with Hadoop	1
2.6	Scaling Out- Hadoop Streaming	1
2.7	Developing a Map Reduce Application-Anatomy of a Map Reduce Job	1
2.8	Scheduling-Shuffle and Sort	1
MODULE 3- Hbase and Pig-8 HRS		
3.1	Introduction to HBase	1
3.2	Filesystems for Hbase	1
3.3	Client API - The Basics	1
3.4	Hbase clients- REST	1
3.5	Shell Commands, Map Reduce Integration	1
3.6	Introduction to Pig-Grunt,	1
3.7	pig data model, Pig Latin, Advanced pig Latin	1
3.8	developing and testing Pig Latin scripts, Map Reduce Integration	1
MODULE 4—ELK- 8 HRS		
4.1	Introduction to ELK, general architecture of ELK,	1
4.2	Elastic search, Installation, and configuration of elastic search, Cluster setup in elastic search,	1
4.3	Indexing, Mapping, Searching	1
4.4	Populate elastic search, Elasticsearch API Conventions,	1
4.5	Elasticsearch Document APIs, Search APIs in Elasticsearch	1
4.6	DSL queries, Analysis in Elasticsearch,	1

4.7	Check the performance of Elasticsearch, API available in elastic search	1
4.8	Create filter, get data and put data, check the health of elastic search	1
MODULE 5-Logstash & Kibana-8 HRS		
5.1	Logstash, Installation, and Configuration of Logstash	1
5.2	Integrating logstash with Kibana, Collection and Parsing of Logs,	1
5.3	log shipping, shipping tools,	1
5.4	filebeat, configure different filebeat, integrate filebeat to Logstash	1
5.5	Kibana, Installation of Kibana,	1
5.6	Interact filebeat, Logstash elastic search, and Kibana,	1
5.7	create an alert in Kibana, docker role in ELK	1
5.8	migrate ELK to Docker	1

Reference Books/ Textbooks:

1. Michael Minelli, Michelle Chambers, and AmbigaDhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses," Wiley,2013

2. P. J. Sadalage and M. Fowler, "NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence", Addison-Wesley Professional,2012.

3. Tom White, "Hadoop: The Definitive Guide", Third Edition, O'Reilley,2012.

4. Asjad Athick, Shay Banon 'Getting Started with Elastic Stack 8.0: Run powerful and scalable data platforms to search, observe, and secure your organization', © 2022 Packt Publishing

5. Vishal Sharma, 'Beginning Elastic Stack' Apress, 2016 ISBN: 978-1-4842-1694-1

File Ref.No.KTU/AC1/804/2021

APJ Abdul Kalam Technological University
Thiruvananthapuram

Abstract

APJAKTU - Academic - M Tech 2022 regulations -Syllabus of Audit courses for the M.Tech Program-recommended by Board of Studies (PG)-sanctioned-Order issued.

ACADEMIC SECTION

U.O.No. 2534/2023/KTU Thiruvananthapuram, Dated: 05.10.2023

Read:-1. Minutes of the meeting of the Syndicate Standing Committee on Academic and Research held on 22.09.2023 in item no.SCAR-046-A01

2. Minutes of the meeting of the Syndicate held on 25.09.2023 in item no.S-049-OA3

3. Minutes of the meeting of the BOS (PG) held on 29.09.2023

4. Approval of the members of SSC A and R via e-mail circulated to member dated:02.10.2023

ORDER

The syllabi for the various audit courses for M.Tech recommended by the Board of Studies in Engineering (PG) were placed before the Syndicate Standing Committee on Academic and Research for making appropriate recomendations to the Syndicate. The SSC on A&R recommended to place the matter in the Syndicate. The SSC on A&R also recommended that the contents in the syllabi for the audit courses shall be reviewed by the BOS (PG) and adequate modifications made vide paper read as 1 above.

The Syndicate of the University vide paper read as 2 resolved that the content in the syllabus for the audit courses shall be reviewed by the BOS (PG) and placed for approval in SSC A&R and thereafter the Syndicate entrusted the Hon'ble Vice Chancellor to approve the audit courses for M.Tech S3 invoking Section 14 (5) of the APJAKTU Act 17 of 2015.

The BOS Engineering (PG) has revised the syllabus and submitted the syllabus for ten audit courses as per the suggestions of the 49th syndicate. The syllabi of various audit courses for M.Tech S3 received from BoS Engineering (PG) were circulated among the members of the SSC on A&R for approval. The members of the SSC on A&R approved the syllabus for the following audit courses:

1. 223AGE100 -ACADEMIC WRITING
2. 223AGE001- ADVANCED ENGINEERING MATERIALS
3. 223AGE003 - DATA SCIENCE FOR ENGINEERS
4. 223AGE004 - DESIGN THINKING
5. 223AGE005 - FUNCTIONAL PROGRAMMING IN HASKELL
6. 223AGE009 - PRINCIPLES OF AUTOMATION
7. 223AGE010 - REUSE AND RECYCLE TECHNOLOGY
8. 223AGE012 - EXPERT SYSTEMS
9. 223AGE011- SYSTEM MODELLING
10. 223AGE002- FORENSIC ENGINEERING

As per the academic calendar published by the university, the M.Tech S3 classes commence on October 13, 2023. The syllabi for audit courses need to be circulated to all colleges, Hence, considering the urgency, sanction is accorded by the Hon'ble Vice Chancellor to approve the syllabi of M.Tech Audit courses recommended by the Board of Studies in Engineering (PG) implementing the Syndicate decision.

This order is issued under Sub Section (5) of Section 14 of the APJ Abdul Kalam Technological University Act, 2015 (17 of 2015).

The syllabi of audit courses for the M.Tech Programme are attached as an annexure to this order.

Orders are issued accordingly.

Sd/-

Dr Saji Gopinath *

Vice Chancellor

Copy to:-

1. The Principals of KTU Affiliated Colleges
2. PS to VC
3. PA to Registrar/ Dean Academic/Controller of Examination
4. JR Administration
5. AR, Academics
6. JD (IT)
7. Director, Academics

8. Joint Director, Academics
9. Chairman ,BOS, PG

Forwarded / By Order

Section Officer

* This is a computer system (Digital File) generated letter. Hence there is no need for a physical signature.

APJ ABDUL KALAM
TECHNOLOGICALUNIVERSITY

(A State Government University)

CET Campus,Thiruvananthapuram, Kerala, INDIA – 695016

Website: https://ktu.edu.in

MTECH AUDIT COURSES

CODE 223AGE100	ACADEMIC WRITING	CATEGORY	L	T	P	CREDIT
		AUDIT COURSE	3	0	0	NIL

Preamble: Learning academic writing sharpens minds, teaches students how to communicate, and develops their thinking capacities and ability to understand others. Writing is thinking, and every student deserves to be a strong thinker. It can also make them think more carefully about what they write. Showing work to others can help to foster a better culture of learning and sharing among students. It also gives students a sense of how they are contributing to the body of work that makes up an academic subject.

Course Outcomes: The COs shown are only indicative. For each course, there can be 4 to 6 COs.

After the completion of the course the student will be able to

CO 1	Understand the principles of scientific/ academic writing
CO 2	Analyse the technique of scientific writing from the reader's perspective
CO 3	Apply the concepts of setting expectations and laying the progression tracks
CO 4	Evaluate the merits of a title, abstract , introduction, conclusion and structuring of a research paper
CO 5	Justify the need using a project proposal or a technical report
CO 6	Prepare a review paper, an extended abstract and a project proposal

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1		3	1				
CO 2		3	1				
CO 3		3	1			2	
CO 4		3	1				
CO 5		3	2	2		2	
CO 6	1	3	3	2		2	

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	40%
Analyse	30%
Evaluate	30%

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern: 40 marks

Course based task : 15 marks
Seminar/Quiz : 15 marks
Test paper, 1 no. : 10 marks
Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern: 60 marks

The examination will be conducted by the respective College. The examination willbe for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any Vive. Each question can carry 12 marks.

Model Question paper

		SET1	**Total Pages:**
Reg No.:________________		Name:____________________________	

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
THIRD SEMESTER M.TECH DEGREE EXAMINATION, MARCH 2024

Course Code: 223AGE100

Course Name: Academic Writing

Max. Marks: 60		Duration: 2.5 Hours

Answer any five full questions, each carries 12 marks.

1 a)	Make clear-cut distinctions between 6 factors that take their toll on readers' memory.	6
1 b)	How can you sustain the attention of the reader to ensure continuous reading?	6
2 a)	What are the different methods by which you can create expectations in the reader?	6
2 b)	Give an account of the topic and non-topic based progression schemes.	6
3 a)	Bring out the differences between an abstract and the introduction of a research paper.	8
3 b)	How are the title of the research paper and its structure related?	4
4	What are 7 principles for including visuals in your research paper. What are the recommended constituents of a conclusion segment of a research paper?	12

5	Give a detailed description of the process and contents of a project proposal for funding.	12
6 a)	What are the contexts recommended for choosing between active and passive voices in technical writing?	8
6 b)	What are the different visual forms that are relevant in a research paper and how do you choose them?	4
7	Give the design of a research paper with the purposes each part serves.	12
	****	

Syllabus and Course Plan (For 3 credit courses, the content can be for 40 hrs and for 2 credit courses, the content can be for 26 hrs. The audit course in third semester can have content for 30 hours).

Syllabus:

CODE 223AG E100	**ACADEMIC WRITING**	**Audit**
Module No.	Topics in a module	Hours
1	Fundamentals of Academic writing from a reader's perspective: acronyms, synonyms, pronouns, disconnected phrases, background ghettos, abusive detailing, cryptic captions, long sentences : all that take their toll on readers' memory.	6
2	Fluid reading & reading energy consumption: setting expectations and laying Progression tracks; Reading energy consumption	6
3	How to write the Title, abstract, introduction ; Structure the writing with headings & subheadings	6
4	Visuals: Resources, Skills, and Methods; Conclusion; References; Bibliography; Grammar in technical writing	6
5	Techniques of writing: An extended abstract, a project proposal, a research paper, a technical report.	6

Course Plan:

No	Topic	No. of Lectures
1	Fundamentals of Academic writing from a reader's perspective: acronyms, synonyms, pronouns, disconnected phrases, background ghettos, abusive detailing, cryptic captions, long sentences all take their toll on readers' memory.	
1.1	The Reading tool-kit to reduce memory required; reduce reading time	1
1.2	Acronyms, Pronouns, Synonyms; Background, broken couple, words overflow	1
1.3	Sustain attention: Keep the story moving forward; Twists, shouts, Pause to clarify, recreate suspense	2

1.4	Keep the reader motivated: Fuel and meet Expectations; Bridge knowledge gap: ground level; Title words; Just In Time to local background	2
2	Fluid reading & reading energy consumption: setting expectations and laying Progression tracks; Reading energy consumption	
2.1	Setting expectations of the reader from Grammar, from theme	1
2.2	Progression tracks for fluid reading: Topic & stress; topic and non topic based progression tracks; pause in progression	2
2.3	Detection of sentence fluidity problems: No expectations/ Betrayed expectations	2
2.4	Controlling reading energy consumption: the energy bill; Energy fuelling stations: Pause	1
3	How to write the Title, abstract, introduction ; Structure the writing with headings & subheadings	
3.1	Title: Face of the paper: Techniques, Qualities & Purpose of title; Metrics	1
3.2	Abstract: Heart of the paper: 4 parts; coherence; tense of verbs, precision; purpose & qualities of the abstract; Metrics	2
3.3	Structure: Headings & sub-headings: Skeleton of the paper: principles for a good structure; Syntactic rules; Quality & Purpose of structures; Metrics	1
3.4	Introduction: Hands of the paper: Start, finish; scope, definitions; answers key reader questions; As a personal active story; Traps, qualities; Metrics	2
4	Visuals: Resources, Skills, and Methods; Conclusion; References; Bibliography; Grammar in technical writing	
4.1	Visuals as the voice of your paper: principles; purpose & qualities of visuals; metrics	2
4.2	Conclusion: contents; purpose, quality; metrics; Abstracts Vs. Conclusion; examples, counter-examples	1
4.3	References, Bibliography: Styles, punctuation marks, quotes, citations	1
4.4	Grammar in Technical writing: Articles, Syntax, Main and subordinate clauses; Active & passive voices; some commonly made mistakes in technical writing.	2
5	Techniques of writing: An extended abstract, a project proposal, a research paper, a technical report.	
5.1	Extended abstract: abstract and keywords, introduction and objective, method, findings and argument, conclusion and suggestions and references.	1
5.2	Project Proposal:Types, executive summary, background including status, objectives, solution, milestones, deliverables, timelines, resources, budgeting, conclusion	2
5.3	Research paper: writing an overview article: provide a comprehensive foundation on a topic; explain the current state of knowledge; identify gaps in existing studies for potential future research; highlight the main methodologies and research techniques	2

5.4	Writing Technical Reports: Title page; Summary; Table of contents; Introduction; Body; Figures, tables, equations and formulae; Conclusion; Recommendations.	1
		30

Reference Books

1. SCIENTIFIC WRITING 2.0 A Reader and Writer's Guide: Jean-Luc Lebrun, World ScientiVic Publishing Co. Pte. Ltd., 2011

2. How to Write and Publish a ScientiVic Paper: Barbara Gastel and Robert A. Day, Greenwood publishers, 2016

3. Grammar, Punctuation, and Capitalisation; a handbook for technical writers and editors. www.sti.nasa.gov/publish/sp7084.pdf www.sti.nasa.gov/sp7084/contents.html

4. Everything You Wanted to Know About Making Tables and Figures. http://abacus.bates.edu/%7Eganderso/biology/resources/writing/ HTWtableVigs.html

223AGE001	ADVANCED ENGINEERING MATERIALS	CATEGORY	L	T	P	CREDIT
		AUDIT COURSE	3	0	0	-

Preamble: This course is designed in a way to provide a general view on typically used advanced classes of engineering materials including metals, polymers, ceramics, and composites.

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

CO 1	Analyse the requirement and find appropriate solution for use of materials.
CO 2	Differentiate the properties of polymers, ceramics and composite materials.
CO 3	Recognize basic concepts and properties of functional materials.
CO 4	Comprehend smart and shape memory materials for various applications.
CO 5	Appraise materials used for high temperature, energy production and storage applications.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	□				□	□	
CO 2	□				□	□	
CO 3	□				□	□	
CO 4	□				□	□	
CO 5	□				□	□	

Assessment Pattern

Bloom's Category	End Semester Examination
Understand	60%
Apply	20%
Analyse	20%

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern: 40 marks

Course based task : 15 marks
Seminar/Quiz : 15 marks
Test paper, 1 no. : 10 marks
Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern: 60 marks

The examination will be conducted by the respective College. The examination willbe for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any Vive. Each question can carry 12 marks.

Model Question paper

AUDIT COURSE

223AGE001 - ADVANCED ENGINEERING MATERIALS

(Answer any five questions. Each question carries 12 Marks)

1.	a) State the relationship between material selection and processing.	**5**
	b) Write about the criteria for selection of materials with respect to the cost and service requirements for engineering applications.	**7**
2.	a) Differentiate thermosetting and thermoplastics with suitable examples.	**5**
	b) Briefly discuss about the properties and applications of polymer nano composite materials.	**7**
3.	a) Write about the potential application areas of functionally graded materials.	**5**
	b) With a neat sketch describe any one processing technique of functionally graded materials.	**7**
4.	a) “Smart materials are functional”? Justify the statement.	**5**
	b) Explain the terms electrostriction and magnetostriction with its application.	**7**

5.	a) What are the factors influencing functional life of components at elevated temperature?	**5**
	b) What are super alloys and what are their advantages?	**7**
6	a) What is a shape memory alloy? What metals exhibit shape memory characteristics?	**4**
	b) Explain about the detection capabilities and uses of pyroelectric sensors.	**8**
7	a) Differentiate between conventional batteries and fuel cells.	**4**
	b) Explain the construction and working of a Li-ion battery.	**8**

Syllabus

Module	**Content**	**Hours**	**Semester Exam Marks (%)**
I	Requirements / needs of advanced materials. Classification of materials, Importance of materials selection, Criteria for selection of materials; motivation for selection, cost basis and service requirements. Relationship between materials selection and processing.	**5**	**20**
II	Classification of non-metallic materials. Polymer, Ceramics: Properties, processing and applications. Nano Composites - Polymer nanocomposites (PNCs), Processing and characterisation techniques – properties and potential applications.	**7**	**20**
III	Functionally graded materials (FGMs), Potential Applications of FGMs, classification of FGMs, processing techniques. limitations of FGMs.	**6**	**20**
IV	Smart Materials: Introduction, smart material types - pyroelectric sensors, piezoelectric materials, electrostrictors and magnetostrictors, shape memory alloys – associated energy stimulus and response forms, applications.	**5**	**20**
V	High Temperature Materials: super alloys – main classes, high temperature properties of superalloys, applications. Energy Materials: materials for batteries.	**7**	**20**

Course Plan

No	Topic	No. of Lectures
1	**Selection of materials for engineering applications**	
1.1	Benefits of advanced materials, classification of materials, importance of materials selection	2
1.2	Selection of materials for different properties, strength, toughness, fatigue and creep	1
1.3	Selection for surface durability, corrosion and wear resistance	1
1.4	Relationship between materials selection and processing	1
2	**Classification of non-metallic materials & nano composites**	
2.1	Rubber: properties, processing and applications.	1
2.2	Plastics: thermosetting and thermoplastics, applications and properties.	2
2.3	Ceramics: properties and applications.	1
2.4	Introduction to nano composites, classification	1
2.5	Processing and characterisation techniques applicable to polymer nanocomposites.	2
3	**Functionally graded materials**	
3.1	General concept, Potential Applications of FGMs	2
3.2	Classification of FGMs	1
3.3	FGMs processing techniques: powder metallurgy route, melt-processing route	2
3.4	Limitations of FGMs	1
4	**Smart materials**	
4.1	Introduction to smart materials, types	1
4.2	Pyroelectric sensors-material class, stimulus, detection capabilities and uses	1
4.3	Piezoelectric materials- material class, stimulus, sensing and actuating applications	1
4.4	Electrostrictors and magnetostrictors - material class, stimulus, micro positioning capabilities and applications	1
4.5	Shape memory alloys (SMAs) - material class, stimulus, temperature sensing and high strain responses, applications.	1
5	**High Temperature Materials and Energy Materials**	
5.1	Characteristics of high-temperature materials, superalloys as high-temperature materials	1
	superalloys - properties and applications	2
5.2	Introduction to lithium-ion battery (LIBs), operating mechanisms and applications	2
5.3	Introduction to Zn-based battery system, types and existing challenges	2

Reference Books

1. DeGarmo et al, "Materials and Processes in Manufacturing", 10th Edition, Wiley, 2008.
2. R.E. Smallman and A.H.W. Ngan, Physical Metallurgy and Advanced Materials, Seventh Edition, Butterworth-Heinemann, 2007
3. Vijayamohanan K. Pillai and Meera Parthasarathy, "Functional Materials: A chemist's perspective", Universities Press Hyderabad (2012).
4. M.V. Gandhi, B.S. Thompson: Smart Materials and Structures, Chapman & Hall, 1992.
5. G. W. Meetham and M. H. Van de Voorde, Materials for High Temperature Engineering Applications (Engineering Materials) Springer; 1 edition (May 19, 2000)
6. Inderjit Chopra, Jayant Sirohi, "Smart Structures Theory", Cambridge University Press, 2013

223AGE003	DATA SCIENCE FOR ENGINEERS	CATEGORY	L	T	P	CREDIT
		AUDIT COURSE	3	0	0	0

Preamble: This course covers essentials of statistics and Linear Algebra and how to prepare the data before processing in real time applications. The students will be able to handle missing data and detection of any outliers available in the dataset. This course explores data science, Python libraries and it also covers the introduction to machine learning for engineers.

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

CO 1	Study Data Science Concepts and statistics
CO 2	Demonstrate Understanding of Mathematical Foundations needed for Data Science
CO 3	Understand Exploratory analysis and Data Visualization and Preprocessing on given dataset
CO 4	Implement Models such as Naive Bayes, K-Nearest Neighbors, Linear and Logistic Regression
CO 5	Build real time data science applications and test use cases

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO7
CO 1	2		2			2	
CO 2	2		2	1		2	
CO 3	2		2	2	2	2	
CO 4	2		2	2	3	2	
CO 5	2		2	3	3	3	2

Assessment Pattern

Bloom's Category	End Semester Examination
Understand	50%
Apply	30%
Analyse	20%

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern: 40 marks

Course based task (Project/Assignments/Simulations/Case studies): 15 marks

Seminar/Quiz : 15 marks

Test paper, 1 no. : 10 marks

Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern:**60 marks**

The examination will be conducted by the respective College. The examination will be for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any five. Each question can carry 12 mark.

Syllabus

Module	Content	Hours	Semester Exam Marks (%)
I	**Statistics for Data science** Probability: Basic concepts of probability, conditional probability, total probability, independent events, Bayes' theorem, random variable, Population, Sample, Population Mean, Sample Mean, Population Distribution, Sample Distribution and sampling Distribution, Mean, Mode, Median, Range, Measure of Dispersion, Variance, Standard Deviation, Gaussian/Normal Distribution, covariance, correlation.	6	20
II	**Linear Algebra** Vectors and their properties, Sum and difference of Vectors, distance between Vectors, Matrices,Inverse of Matrix, Determinant of Matrix, Trace of a Matrix, Dot Product, Eigen Values, Eigen Vectors, Single Value Decomposition	6	20
III	**Hypothesis Testing** Understanding Hypothesis Testing, Null and Alternate Hypothesis, Non-directional Hypothesis, Directional Hypothesis Critical Value Method, P-Value Method, Types of Errors-Type1 Error,Type2 Error, Types of Hypothesis Test Z Test, Chi-Square	6	20

IV	**Exploratory Data Analysis** Data Collection –Public and Private Data, Data Cleaning-Fixing Rows and Columns, Missing Values, Standardizing values, invalid values, filtering data, Data-Integration,Data-Reduction,Data Transformation	**6**	**20**
V	**Machine Learning and Python for Data Science** Python Data structures-List, Tuple, Set, Dictionary, Pandas, Numpy, Scipy, Matplotlib, Machine Learning-Supervised Machine Learning, Unsupervised Machine Learning,Regression, Classification, Naïve-Bayes	**6**	**20**

Course Plan

No	Topic	No. of Lectures
1	**Statistics for Data science**	
1.1	Probability: Basic concepts of probability, conditional probability, total probability	1
1.2	independent events, Bayes' theorem, random variable, Population	1
1.3	Sample, Population Mean, Sample Mean, Population Distribution	1
1.4	Sample Distribution and sampling Distribution, Mean, Mode, Median, Range, Propositional logic and predicate logic	1
1.5	Measure of Dispersion, Variance, Standard Deviation	1
1.6	Gaussian/Normal Distribution, covariance, correlation.	1
2	**Linear Algebra**	
2.1	Vectors and their properties,	1
2.2	Sum and difference of Vectors, distance between Vectors	1
2.3	Matrices,Inverse of Matrix,	2
2.4	Determinant of Matrix, Trace of a Matrix, Dot Product, Eigen Values, Eigen Vectors, Single Value Decomposition	2
3	**Hypothesis Testing**	
3.1	Understanding Hypothesis Testing, Null and Alternate Hypothesis	1
3.2	Non-directional Hypothesis, Directional Hypothesis Critical Value Method, P-Value Method,	2
3.3	Types of Errors-Type1 Error,Type2 Error,	1
3.4	Types of Hypothesis Test Z Test, Chi-Square,	2
4	**Exploratory Data Analysis**	
4.1	Data Collection –Public and Private Data	1
4.2	Data Cleaning-Fixing Rows and Columns	1
4.3	Missing Values	1
4.4	Standardizing values	1
4.5	Invalid values, filtering data	1
4.6	Data Integration, Data Reduction, Data Transformation	1

5	**Machine Learning and Python for Data Science**	
5.1	Python Data structures-List, Tuple, Set,	1
5.2	Dictionary, Pandas, Numpy, Matplotlib	2
5.3	Machine Learning-Supervised Machine Learning, Unsupervised Machine Learning	1
5.4	Regression, Classification	1
5.5	Naïve-Bayes	1

Reference Books

1. Python Data Science Handbook. Essential Tools for Working with Data, Author(s): Jake VanderPlas, Publisher: O'Reilly Media, Year: 2016
2. Practical Statistics for Data Scientists: 50 Essential Concepts, Author(s): Peter Bruce, Andrew Bruce, Publisher: O’Reilly Media, Year: 2017
3. Practical Linear Algebra for Data Science, by Mike X Cohen, Released September 2022, Publisher(s): O'Reilly Media, Inc.
4. Data Science from Scratch ‘by Joel Grus, Released, April 2015, Publisher(s): O'Reilly Media, Inc.
5. Hands-On Exploratory Data Analysis with Python, by Suresh Kumar Mukhiya, Usman Ahmed, Released March 2020,Publisher(s): Packt Publishing

SET1 **Total Pages:**

Reg No.:_ _ Name:____________________________

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

THIRD SEMESTER M.TECH DEGREE EXAMINATION, MARCH 2024

Course Code: 223AGE003

Course Name: DATA SCIENCE FOR ENGINEERS

Max. Marks: 60 Duration: 2.5 Hours

Answer any five full questions, each carries 12 marks.

1 a) It is observed that 50% of mails are spam. There is software that filters spam mail before reaching the inbox. It accuracy for detecting a spam mail is 99% and chances of tagging a non-spam mail as spam mail is 5%. If a certain mail is tagged as spam finds the probability that it is not a spam mail. 5

b) Depict the relevance of measures of central tendency in data wrangling with a suitable example 7

2. a) Calculate the inverse of the Matrix 4

$$\begin{matrix} 2 & 4 & -6 \\ 7 & 3 & 5 \\ 1 & -2 & 4 \end{matrix}$$

b)Find all Eigenvalues and Corresponding Eigenvectors for the matrix if 8

$$\begin{matrix} 2 & -3 & 0 \\ 2 & -5 & 0 \\ 0 & 0 & 3 \end{matrix}$$

3. a) A statistician wants to test the hypothesis H0: $\mu = 120$ using the alternative hypothesis Hα: $\mu > 120$ and assuming that $\alpha = 0.05$. For that, he took the sample values as $n = 40$, $\sigma = 32.17$ and $\bar{x} = 105.37$. Determine the conclusion for this hypothesis? 5

b) Hypothesis testing is an integral part of statistical inference, list out the various types of hypothesis testing and also mentions their significances in data science. 7

4. a) Brief in detail directional and non-directional hypothesis 6

b) Differentiate null and alternate hypothesis and also elaborate on type 1 and type 2 errors 6

5. a) Explain the concepts of Tuple, List and Directory in python with example 6

b) Elucidate reinforcement learning and application in real world. 6

6. a) What is Feature Engineering , demonstrate with an example 6

b) Describe in detail different steps involved in data preprocessing. 6

7. a) Illustrate supervised learning model with linear regression model 5

b) Predict the probability for the given feature vector if an accident will happen or not? 7

Weather condition: rain, Road condition: good, Traffic condition: normal, Engine problem: no, the task is to predict using Naïve Bayes classification.

SNo.	Weather condition	Road condition	Traffic condition	Engine problem	Accident
1	Rain	bad	high	no	yes
2	snow	average	normal	yes	yes
3	clear	bad	light	no	no
4	clear	good	light	yes	yes
5	snow	good	normal	no	no
6	rain	average	light	no	no
7	rain	good	normal	no	no
8	snow	bad	high	no	yes
9	clear	good	high	yes	no
10	clear	bad	high	yes	yes

223AGE004	DESIGN THINKING	CATEGORY	L	T	P	CREDIT
		AUDIT COURSE	3	0	0	-

Preamble:

This course offers an introductory exploration of fundamental engineering concepts and techniques, the design process, analytical thinking and creativity, as well as the fundamentals and development of engineering drawings, along with their application in engineering problems.

Course Outcomes:

After the completion of the course the student will be able to

CO 1	Identify and frame design challenges effectively.
CO 2	Generate creative ideas through brainstorming and ideation
CO 3	Iterate on designs based on user insights
CO 4	Apply Design Thinking to real-world problems and projects.

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1				2		2	2
CO 2	2		2	2			2
CO 3		2		2		2	2
CO 4	2		2	3	2		2

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	40
Analyse	30
Evaluate	30
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

AUDIT COURSES

Continuous Internal Evaluation Pattern: 40 marks

Course based task : 15 marks
Seminar/Quiz : 15 marks
Test paper, 1 no. : 10 marks
Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern: 60 marks

The examination will be conducted by the respective College. The examination willbe for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any five. Each question can carry 12 marks.

Model Question paper

		SET1	**Total Pages:**
Reg No.:_______________		Name:___________________________	
APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY THIRD SEMESTER M.TECH DEGREE EXAMINATION, MARCH 2024			
Course Code: 223AGE004			
Course Name: DESIGN THINKING			
Max. Marks: 60		Duration: 2.5 Hours	
Answer any five full questions, each carries 12 marks.			

1 a)	How can a multidisciplinary team collaborate effectively to implement design principles?	7
1 b)	What are the key differences between human-centred design and other design methodologies?	5
2 a)	How do you measure the success of a design project in terms of user satisfaction and impact?	7
2 b)	How does the iterative nature of the design process contribute to better outcomes	5

3 a)	What are the fundamental principles of effective brainstorming, and how do they differ from traditional problem-solving approaches?	7
3 b)	What are some key principles of ergonomic design, and how do they contribute to the usability and comfort of products?	5
4 a)	Enumerate some examples of successful and unsuccessful market testing scenarios, and what lessons can be learned from these experiences to improve future product or service launches?	7
4b)	What is the primary purpose of creating prototypes in the design and development process?	5
5	What strategies and methodologies can designers use to embrace agility and respond quickly to changing user needs and market dynamics?	12
6	Illustrate any four examples of successful bio-mimicry applications in various industries.	12
7	What ethical considerations should designers keep in mind when designing for diverse user groups?	12
****		

Syllabus:

Module 1

Design process: Traditional design, Design Thinking Approach, Introduction to Design Thinking, History and evolution of Design Thinking, Role of design thinking in the human-centred design process. Design space, Design Thinking in a Team Environment, Team formation.

Module 2

Design Thinking Stages: Empathize, Define, Ideate, Prototype and Test. The importance of empathy, Building a user-centred mindset. Problem statement formulation, User needs and pain points, establishing target specifications, Setting the final specifications.

Module 3

Generating Ideas, Brainstorming techniques, Application of Aesthetics and Ergonomics in Design. Bio-mimicry, Conceptualization, Visual thinking, Drawing/Sketching, Presenting ideas.

Module 4

Use of prototyping, Types of prototypes, Rapid prototyping techniques, User testing and feedback collection, Iterative prototyping, testing to gauge risk and market interest

Module 5

Entrepreneurship/business ideas, Patents and Intellectual Property, Agility in design, Ethical considerations in design. Overcoming common implementation challenges

Corse Plan SyllabusandCorsePlan (For 3credit courses, thec ontent can be for 40 hrs and for2credit courses, the content can be for 26 hrs. The audit course in third semester can have content for 30hours).

No	Topic	No. of lectures
1	**Design process:**	
1.1	Design process: Traditional design, Design Thinking Approach, Introduction to Design Thinking, History and evolution of Design Thinking.	3
1.2	Role of design thinking in the human-centred design process. Design space,	2
1.3	Design Thinking in a Team Environment, Team formation.	2

2	**Design Thinking Stages:**	
2.1	Design Thinking Stages: Empathize, Define, Ideate, Prototype and Test.	2
2.2	The importance of empathy, Building a user-centred mindset.	2
2.3	Problem statement formulation, User needs and pain points, establishing target specifications, Setting the final specifications.	3
3	**Ideation**	
3.1	Generating Ideas, Brainstorming techniques.	2
3.2	Application of Aesthetics and Ergonomics in Design. Bio-mimicry.	3
3.3	Conceptualization, Visual thinking, Drawing/Sketching, Presenting ideas.	2
4	**Prototyping and testing**	
4.1	Use of prototyping, Types of prototypes, Rapid prototyping techniques.	3
4.2	User testing and feedback collection, Iterative prototyping, testing to gauge risk and market interest	2
5	**IPR in design**	
5.1	Entrepreneurship/business ideas, Patents and Intellectual Property.	2
5.2	Agility in design, Ethical considerations in design. Overcoming common implementation challenges	2

Reference Books

1. Christoph Meinel, Larry Leifer and Hasso Plattner- "Design Thinking: Understand – Improve – Apply", Springer Berlin, Heidelberg, 2011.
2. Thomas Lockwood and Edgar Papke – "Design Thinking: Integrating Innovation, Customer Experience, and Brand Value", Allworth Press, 2009.
3. Pavan Soni – "Design Your Thinking", Penguin Random House India Private Limited, 2020.
4. Andrew Pressman- "Design Thinking : A Guide to Creative Problem Solving for Everyone", Taylor & Francis, 2018.
5. N Siva Prasad, "Design Thinking Techniques an Approaches" Ane Books Pvt. Ltd.,2023

SYLLABUS

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
223AGE005	FUNCTIONAL PROGRAMMING IN HASKELL	AUDIT COURSE	3	0	0	-

Preamble: This course introduces a functional programming approach in problem solving. Salient features of functional programming like recursion, pattern matching, higher order functions etc. and the implementation in Haskell are discussed.

Course Outcomes:

After the completion of the course the student will be able to

CO 1	Understand the functional programming paradigm which is based on the mathematics of lambda calculus.
CO 2	Develop Haskell programs using functions, guards and recursive functions
CO 3	Apply the concept of tuples, lists and strings in Haskell programming
CO 4	Apply the concept of algebraic data types, abstract data types, modules, recursive data types and user defined data types in Haskell programming
CO 5	Develop Haskell programs with files for reading input and storing output

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1					3		
CO 2	2			2	3		
CO 3	2			2	3		
CO 4	2			2	3		
CO 5	2			2	3		

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	40%
Analyse	40%
Evaluate	20%
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation: 40 marks

Course based task	: 15 marks
Seminar/Quiz	: 15 marks
Test paper, 1 no.	: 10 marks

Test paper shall include minimum 80% of the syllabus**.**

End Semester Examination: 60 marks

The examination will be conducted by the respective College. The examination will be for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any five. Each question can carry 12 marks.

Model Question paper

		Total Pages:
Reg No.:_______________		Name:___________________________

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
THIRD SEMESTER M.TECH DEGREE EXAMINATION, DECEMBER 2023

Course Code: 223AGE005

Course Name: Functional Programming in Haskell

Max. Marks: 60		Duration: 2.5 Hours

Answer any five full questions, each carries 12 marks.

1 a.	Explain the basic differences between imperative style programming and functional style programming.	3
1 b.	Analyse each of the following lambda expressions to clarify its structure. If the expression is a function, identify the bound variable and the body expression, and then analyse the body expression. If the expression is an application, identify the function and argument expressions, and then analyse the function and argument expressions: i) λa.(a λb.(b a)) ii) λx.λy.λz.((z x) (z y)) iii) (λf.λg.(λh.(g h) f) λp.λq.p)	9
2 a.	Design a recursive function to find 2^n where n is a natural number.	4

2 b.	Explain various forms of function definitions in Haskell with the help of examples.	8
3 a.	Explain any three list operations along with function definitions and examples.	6
3 b.	Write a program to duplicate only even numbers among the elements of a list using a Haskell function by (i) Recursion (ii) List Comprehension and explain. Example : λ> dupli [1, 2, 3] ANS: [2,2]	6
4	Write Recursive definitions along with an explanation for the below arithmetic operations. Illustrate the recursive flow with the help of a diagram. i. add x y ii. mult x y iii. div x y	12
5	Write the Haskell code to split a list into two lists such that the elements with odd index are in one list while the elements with even index are in the other list.	12
6 a	Give the type definition of a binary tree along with explanation of two functions on binary trees.	6
6 b	Define a queue data type in Haskell along with any two operations on it with examples.	6
7 a.	Explain the basic steps of reading from files and writing to files in Haskell.	4
7 b.	Write a Haskell program to read from the file "input.txt", display the contents on the screen and write the contents to another file "output.txt".	8
****		

Syllabus and Corse Plan (For 3 credit courses, the content can be for 40 hrs and for 2 credit courses, the content can be for 26 hrs. The audit course in third semester can have content for 30 hours).

Module 1 (5 Hrs)

Introduction to Functional Programming: Programming language paradigms, imperative style programming, comparison of programming paradigms.
Functional programming, Functions - Mathematical concepts and terminology, Lambda calculus, Function definitions, programs as functions, Functional programming Languages. Haskell basics, GHCi interpreter.

Module 2 (6 Hrs)

Programming in Haskell: Expressions and evaluation, Lazy evaluation, let expressions, scopes.

Basic data types in Haskell, operators, infix operators, associativity and precedence, Arithmetic functions.

types, definitions, currying and uncurrying, type abstraction.

Function definitions, pattern matching, guards, anonymous functions, higher order functions.

Recursion, Programming exercises.

Module 3 (7 Hrs)

Data types: tuples and lists: Tuples , Lists: building lists, decomposing lists, functions on lists, built-in functions on lists, primitive and general recursion over lists, infinite lists.

Strings: functions on strings.

Polymorphism and overloading, conditional polymorphism

Module 4 (6 Hrs)

Type classes, Algebraic data types, Modules, Recursive data types.

User defined data types, Records, Stacks, Queues, Binary trees, Constructors, Destructors.

Module 5 (6 Hrs)

Functor, Applicative functor, Monad

Programming with actions: Functions vs actions, Basics of input / output, the do notation, interacting with the command line and lazy I/O, File I/O.

No	Topic	No. of Lectures
1	**Introduction to Functional Programming**	
1.1	Programming language paradigms, imperative style programming, comparison of programming paradigms	1
1.2	Functional programming, Functions - Mathematical concepts and terminology	1
1.3	Lambda calculus	1
1.4	Function definitions, programs as functions, Functional programming Languages	1
1.5	Haskell basics, GHCi interpreter	1
2	**Haskell basics**	
2.1	Expressions and evaluation, Lazy evaluation	1
2.2	let expressions, scopes, Basic data types in Haskell	1
2.3	operators, infix operators, associativity and precedence, Arithmetic	1

	functions	
2.4	types, definitions, currying and uncurrying, type abstraction.	1
2.5	Function definitions, pattern matching, Guards	1
2.6	anonymous functions, higher order functions, Recursion	1
3	**Data types: tuples and lists**	
3.1	Tuples , Lists: building lists, decomposing lists	1
3.2	functions on lists, built-in functions on lists	1
3.3	primitive and general recursion over lists	1
3.4	infinite lists	1
3.5	Strings: functions on strings	1
3.6	Polymorphism and overloading	1
3.7	conditional polymorphism	1
4	**User defined data types**	
4.1	Type classes, Algebraic data types, Modules	1
4.2	Recursive data types	1
4.3	User defined data types, Records	1
4.4	Stacks, Queues	1
4.5	Binary trees	1
4.6	Constructors, Destructors	1
5	**Programming with actions**	
5.1	Functor, Applicative functor,	1
5.2	Monad	1
5.3	Programming with actions: Functions vs actions, Basics of input / output, the do notation	1
5.4	interacting with the command line and lazy I/O	1
5.5	File I/O	2

Reference Books

[1] Richard Bird, “Introduction to functional programming using Haskell', second edition, Prentice hall series in computer science

[2] Bryan O'Sullivan, Don Stewart, and John Goerzen, “Real World Haskell”

[3] Richard Bird, “Thinking Functionally with Haskell”, Cambridge University Press, 2014

[4] Simon Thompson, “Haskell: The Craft of Functional Programming”, Addison-Wesley, 3rd Edition, 2011

[5] H. Conrad Cunningham, “Notes on Functional Programming with Haskell”, 2014

[6] Graham Hutton, “Programming in Haskell”, Cambridge University Press, 2nd Edition, 2016

[7] Alejandro Serrano Mena, “Practical Haskell: A Real-World Guide to Functional Programming”, 3rd Edition, Apress, 2022

[8] Miran Lipovaca, “Learn You a Haskell for Great Good!: A Beginner's Guide”, No Starch Press, 2011

223AGE010	REUSE AND RECYCLE TECHNOLOGY	CATEGORY	L	T	P	CREDIT
		AUDIT COURSE	3	0	0	-

Preamble: "Reuse and Recycle Technology" typically focuses on sustainable practices and technologies aimed at reducing waste, conserving resources, and promoting environmental responsibility.

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

CO 1	Explain the principles and technologies behind waste reduction, resource conservation, and sustainable practices
CO 2	Describe and Analyze waste generation and management.
CO 3	Apply the knowledge of various reuse strategies and their application in different industries and Analyze various recycling technologies
CO 4	Appraise the methods of E-waste management and Eco friendly packaging
CO 5	Comprehend Environmental Regulations and Policies**,** Understand the importance of environmental regulations and policies in addressing environmental challenges

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6
CO 1			3			
CO 2				3		
CO 3				3		
CO 4					3	
CO 5			3			

Assessment Pattern

Bloom's Category	End Semester Examination
Understand	60%
Apply	20%
Analyse	20%

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern: 40 marks

Course based task : 15 marks
Seminar/Quiz : 15 marks
Test paper, 1 no. : 10 marks
Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern: 60 marks

The examination will be conducted by the respective College. The examination willbe for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any five. Each question can carry 12 marks.

Model Question paper

AUDIT COURSE

223AGE010 - REUSE AND RECYCLE TECHNOLOGY

Answer any five full questions, each carries 12 marks.

1.	(a) What are the 3 pillars of sustainability? *(b)* What is sustainable waste management? What makes sustainable waste management so important?	5 7
2.	(a)How do the three categories of municipal solid waste differ? *(b)* Discuss the municipal waste collection and management?	5 7
3.	(a)Explain the major differences between Reuse and Recycle? *(b)* Give an overview of recycling technologies used for any two materials. Discuss the Process involved.	5 7
4.	(a)What are the common source of E-waste *(b)* What are the challenges and opportunities in E-waste management	5 7
5.	(a)What is the case law for waste recycling in India *(b)* Discuss sustainable packaging and its environmental impacts	5 7
6.	Explain the various environmental regulations in India for addressing Environmental challenges	12
7.	a) Give examples of water reuse technologies in circular economy b) How can we reduce e-waste with sustainable solutions	5 7

Syllabus

Module	Content	Hours	Semester Exam Marks (%)
I	**Introduction to Sustainability**, Understanding sustainability and its importance, The three pillars of sustainability: Environmental, Social, and Economic. Biodiversity conservation, Climate change and mitigation Sustainable resource management.	**6**	**20**
II	**Waste Management,** Definition and classification of waste, Waste Generation and Composition**,** Waste Collection and Transportation, Waste Segregation and Sorting. Waste Disposal Methods Historical perspectives on waste management, The three Rs: Reduce, Reuse, and Recycle.	**6**	**20**
III	**Recycling and Reuse:** Importance of reuse, Application of reuse in various industries, Challenges and opportunities in reuse, Overview of recycling technologies, Circular economy, Sorting and processing of recyclable materials, Advanced recycling methods. Emerging technologies in recycling.	**6**	**20**
IV	**E-waste Recycling,** Challenges and environmental impact of electronic waste, E-waste recycling methods and regulations, Sustainable electronics design, **Sustainable Packaging,** Packaging materials and their environmental impact, Eco-friendly packaging alternatives, Packaging design for sustainability	**6**	**20**
V	**Environmental Regulations and Policies,** Understand the importance of environmental regulations and policies in addressing environmental challenges, National and international waste and recycling regulations, Compliance and enforcement, Industry standards and certifications	**6**	**20**

Course Plan

No	Topic	No. of Lectures
1	**Introduction to Sustainability (6)**	
1.1	Understanding sustainability and its importance	1
1.2	The three pillars of sustainability: Environmental, Social, and Economic.	3
1.3	Biodiversity conservation, Climate change and mitigation	1
1.4	Sustainable resource management	1
2	**Waste Management (6)**	
2.1	Definition and classification of waste	1
2.2	Waste Generation and Composition	1
2.3	Waste Collection and Transportation.	1
2.4	Waste Segregation and Sorting.	1
2.5	Waste Disposal Methods	1
2.6	Historical perspectives on waste management, The three Rs: Reduce, Reuse, and Recycle.	1
3	**Recycling and Reuse (6)**	
3.1	Importance of reuse, Examples of reuse in various industries.	1
3.2	Challenges and opportunities in reuse	1
3.3	Overview of recycling technologies, Sorting and processing of recyclable materials	2
3.4	Advanced recycling methods	1
3.5	Emerging technologies in recycling.	1
4	**E-waste Recycling (6)**	
4.1	Challenges and environmental impact of electronic waste	1
4.2	E-waste recycling methods and regulations	1
4.3	Sustainable electronics design	1
4.4	Packaging materials and their environmental impact	1
4.5	Eco-friendly packaging alternatives	1
4.6	Packaging design for sustainability	1
5	**Environmental Regulations and Policies (6)**	
5.1	Importance of environmental regulations and policies in addressing environmental challenges	2
5.2	National and international waste and recycling regulations	2
5.3	Industry standards and certifications, Compliance and enforcement	2

Reference Books

1. Sustainable Engineering: Concepts, Design and Case Studies, David T. Allen, Pearson Publication.
2. A Comprehensive Book on Solid Waste Management with Application, Dr. H.S. Bhatia , Misha Books, 2019
3. "Cradle to Cradle: Remaking the Way We Make Things" by William McDonough and Michael Braungart.
4. "Recycling of Plastic Materials" edited by Vijay Kumar Thakur
5. E-waste: Implications, Regulations and Management in India and Current Global Best Practices, Rakesh Johri, TERI
6. "Sustainable Packaging", Subramanian Senthilkannan Muthu , Springer Nature.
7. **I**ndian Environmental Law: Key Concepts and Principles " Orient Black swan Private Limited, New Delhi.

223AGE012	EXPERT SYSTEMS	CATEGORY	L	T	P	CREDIT
		AUDIT COURSE	3	0	0	-

Preamble: The course aims to provide an understanding of the basic concepts of Artificial Intelligence (AI) and Expert Systems. The course also covers the knowledge representation in expert systems, classes of expert systems, applications of expert systems.

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

CO 1	Explain the concepts of Artificial Intelligence and different ways of knowledge representations.
CO 2	Explain the components of expert systems, development stages of expert systems and tools available for expert system design.
CO 3	Apply the concept of knowledge representation in expert systems
CO 4	Differentiate the classes of expert systems and examine properties of existing systems

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO7
CO 1	1		2	1	2	2	
CO 2	1		1	3	2	2	
CO 3	1		1	2	2	2	
CO 4	2		2	2	3	2	

Assessment Pattern

Bloom's Category	End Semester Examination
Understand	60%
Apply	20%
Analyse	20%

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern: 40 marks

Course based task (Project/Assignments/Simulations/Case studies): 15 marks

Seminar/Quiz : 15 marks

Test paper, 1 no. : 10 marks

Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern:**60 marks**

The examination will be conducted by the respective College. The examination will be for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any five. Each question can carry 12 mark.

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY THIRD SEMESTER M.TECH DEGREE EXAMINATION, MARCH 2024		
Course Code: 223AGE012		
Course Name: EXPERT SYSTEMS		
Max. Marks: 60		Duration: 2.5 Hours
Answer any five full questions, each carries 12 marks.		
1	a) What are the types of AI? Explain with examples .	6
	b) What do you mean by knowledge in AI and explain the different ways of knowledge representation used in AI?	6
2.	a) Write note on semantic network.	6
	b) What are Predicates? Explain its syntax and semantics.	6
3.	a) Write notes on different tools available for expert system design.	6
	b). What are the different stages in the development of an expert system?	6
4.	a) Illustrate Conceptual Dependencies with an example.	6
	b) Illustrate with an example the Structured Knowledge representation of an Expert System.	6
5.	a) What do you mean by Frame based Expert System? Explain	6
	b)Explain the architecture of MYCIN	6
6.	a)Explain Fuzzy based expert systems	6
	b) Explain the neural network based expert systems	6
7.	a) Explain any two applications of expert systems?	6
	b)What are the limitations of expert system ? Explain	6

Syllabus

Module	Content	Hours	Semester Exam Marks (%)
I	Overview of Artificial Intelligence (AI): Definition & Importance of AI. Knowledge general concepts: Definition and Importance of knowledge, Knowledge-Based Systems, Knowledge organization, Knowledge Manipulation and acquisition. Knowledge Representation: Introduction, Syntax and Semantics- Propositional logic and predicate logic.	6	20
II	Basic concepts of expert systems-Introduction to expert systems, Components of expert systems. Features of Expert System, Stages in the development of expert system, Types of tools available for expert system design	6	20
III	Knowledge representation in expert systems: Structured Knowledge representation: Graphs, Frames and related structures, Associative networks, Conceptual dependencies, Examples of structured knowledge representation.	6	20
IV	Classes of expert systems: Rule-based expert systems, Example- MYCIN, Frame-based expert system, terminologies, IF-THEN structure. Fuzzy and Neural network based expert systems(basic concepts)	7	20
V	Currents trends in expert systems, Advantages and limitations of expert systems, Applications of expert systems.	5	20

Course Plan

No	Topics	No. of Lectures
1	**Overview of Artificial Intelligence& Knowledge general concepts**	
1.1	Definition & Importance of AI	1
1.2	Definition and Importance of Knowledge,	1
1.3	Knowledge-Based Systems, Knowledge Organization	1
1.4	Knowledge Manipulation and acquisition	1
1.5	Knowledge Representation: Introduction, Syntax and Semantics	1
1.6	Propositional logic and predicate logic	1
2	**Basic concepts of expert systems**	
2.1	Introduction to Expert System, Components of expert systems	2
2.2	Features of Expert System, Stages in the development of expert system	2
2.3	Types of tools available for expert system design	2
3	**Knowledge representation in expert systems**	
3.1	Structured Knowledge representation	1
3.2	Graphs, Frames and Related Structures	2
3.3	Associative Networks, Conceptual Dependencies	2
3.4	Examples of structured knowledge representation	1
4	**Classes of expert systems**	
4.1	A rule-based expert system -Introduction	1
4.2	MYCIN	1
4.3	IF-THEN structure	1
4.4	Frame-based expert system	2
4.5	Fuzzy based expert systems	1
4.6	Neural network based expert systems	1
5	**Currents trends and applications of expert systems**	
5.1	Currents trends of expert systems	2
5.2	Advantages and limitations of expert systems	1
5.3	Applications of expert systems	2

Reference Books

1. E. Rich & K. Knight - Artificial Intelligence, 2/e, TMH, New Delhi, 2005.
2. P.H. Winston - Artificial Intelligence, 3/e, Pearson Edition, New Delhi, 2006.
3. D.W. Rolston - Principles of AI & Expert System Development, TMH, New Delhi
4. Kevin Night and Elaine Rich, Nair B., “Artificial Intelligence (SIE) “, McGraw Hill – 2010
5. Dan W Patterson, ‘Introduction to Artificial intelligence and Expert systems’, Prentice Hall of India Pvt. Ltd,2007
6. Russel (Stuart), ‘Artificial Intelligence- Modern approach, Pearson Education series in AI’, 3rd Edition, 2009.
7. I. Gupta, G. Nagpal · Artificial Intelligence and Expert Systems, Mercury Learning and Information -2020

223AGE011	SYSTEM MODELLING	CATEGORY	L	T	P	CREDIT
		AUDIT COURSE	3	0	0	-

Preamble: Study of this course provides the learners a clear understanding of fundamental concepts in simulation and modelling. This course covers the different statistical models, importance of data collection and various types of simulations. The course helps the learners to find varied applications in engineering, medicine and bio-technology.

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

CO 1	Analyse the requirement and find appropriate tool for simulation.
CO 2	Differentiate the different statistical models.
CO 3	Discuss the different techniques for generating random numbers.
CO 4	Analyse the different methods for selecting the different input models..
CO 5	Discuss the different measures of performance and their estimation

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6
CO 1	2		1	1	2	
CO 2	2		1	1	1	
CO 3	1					
CO 4	1		1	1		
CO 5	2		1	1	1	

Assessment Pattern

Bloom's Category	End Semester Examination
Understand	60%
Apply	20%
Analyse	20%

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Course based task (Project/Assignments/Simulations/Case studies): 15 marks
Seminar/Quiz: 15 marks
Test paper, 1 no.: 10 marks
Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern:

The examination will be conducted by the respective College. The examination will be for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any five. Each question can carry 12 marks.

Model Question paper

AUDIT COURSE

223AGE001 – SYSTEM MODELLING

Answer any five questions Each carries 12 marks

PART A

1. a. Discuss the advantages and disadvantages of simulation. (5marks)

   b. What are the areas of applications of simulation (7 marks)

2. a.A bus arrives every 20 minutes at a specified stop beginning at 6:40 A.M. and continuing until 8:40 A.M. A certain passenger does not know the schedule, but arrives randomly (uniformly distributed) between 7:00A.M. and 7:30 A.M. every morning. What is the probability that the passenger waits more than 5 minutes for a bus? (5 marks)

   b. A production process manufactures computer chips on the average at 2% nonconforming. Every day, a random sample of size 50 is taken from the process. If the sample contains more than two nonconforming chips, the process will be stopped. Compute the probability that the process is stopped by the sampling scheme. (7 marks)

3. a.Discuss the different types of tests for random numbers. (5 marks)

   b. Generate random numbers using multiplicative congruential method with X0 = 5, a 11, and m = 64. (7 marks)

4. a. What are the different methods of data collection. (4marks)

   b. Records pertaining to the monthly number of job-related injuries at an underground coalrnine were being studied by a federal agency. The values for the past 100 months were as follows:

Injuries per Month	Frequency of Occurrence
0	35
1	40
2	13
3	6
4	4
5	1
6	1

(a) Apply the chi-square test to these data to test the hypothesis that the underlying distribution is Poisson. Use the level of significance α== 0.05.
(b) Apply the chi-square test to these data to test the hypothesis that the distribution is Poisson with mean 1.0. Again let α = 0.05. ·
c) What are the differences between parts (a) and (b), and when might each case arise? (8 marks)

5. a.What is the difference between validation and verification.(5 marks)
   b. Discuss the different measures of performance and their estimation(7 marks)
6. a. Discuss the different methods of parameter estimation(5 marks)
   b. With an example, describe the Poisson process.(7 marks)
7. a. Distinguish between discrete and continuous systems(5 marks)
   b.What are the different components of a simulation system(7 marks)

Syllabus

Module	Content	Hours	Semester Exam Marks (%)
I	When simulation is the appropriate tool. Advantages and disadvantages of Simulation; Areas of application, Systems and system environment; Components of a system; Discrete and continuous systems, Model of a system; Types of Models, Discrete-Event System Simulation, Steps of a simulation study.	6	20
II	Review of terminology and concepts, Useful statistical models, Discrete distributions. Continuous distributions, Poisson process, Empirical distributions. (basic idea only)	6	20
III	Properties of random numbers; Generation of pseudo-random numbers, Techniques for generating random numbers, Tests for Random Numbers	6	20
IV	Data Collection; Identifying the distribution with data, Parameter estimation, Goodness of Fit Tests, Fitting a non-stationary Poisson process, Selecting input models without data, Multivariate and Time-Series input models.	6	20
V	Measures of performance and their estimation, Output analysis for terminating simulations, Output analysis for steady-state simulations, Verification, calibration and validation	6	20

Course Plan

No	Topic	No. of Lectures
1	**Introduction**	
1.1	When simulation is the appropriate tool	1
1.2	Advantages and disadvantages of Simulation;	1
1.3	Areas of application, Systems and system environment;	1
1.4	Components of a system; Discrete and continuous systems,	1
1.5	Model of a system; Types of Models,	1
1.6	Discrete-Event System Simulation ,Steps of a simulation study	1
2	**Statistical Models in Simulation**	
2.1	Review of terminology and concepts, Empirical distributions. (basic idea only)	1
2.2	Useful statistical models,	1
2.3	Discrete distributions.	1
2.4	Continuous distributions,.	1
2.5	Poisson process	1
2.6	Empirical distributions	1
3	**Random Number Generation**	
3.1	Properties of random numbers;	1
3.2	Generation of pseudo-random numbers,	
3.3	Techniques for generating random numbers	1
3.4	Techniques for generating random numbers(cont)	1
3.5	Tests for Random Numbers	1
3.6	Tests for Random Numbers(cont)	1
4	**Input Modelling**	
4.1	Data Collection;	1
4.2	Identifying the distribution with data.	1
4.3	Parameter estimation, Goodness of Fit Tests	1
4.4	Fitting a non-stationary Poisson process	1
4.5	Selecting input models without data,	1
4.6	Multivariate and Time-Series input models	1
5	**Measures of Performance and their Estimation**	
5.1	Measures of performance and their estimation	1
5.2	Measures of performance and their estimation(cont)	1
5.3	Output analysis for terminating simulations	1
5.4	Output analysis for steady-state simulations	1
5.5	Verification, calibration and validation	1
5.6	Verification, calibration and validation(cont)	1

Textbooks:

1. Jerry Banks, John S. Carson II, Barry L. Nelson, David M. Nicol: Discrete-Event System Simulation, 5th Edition, Pearson Education, 2010.

Reference Books:

1. Lawrence M. Leemis, Stephen K. Park: Discrete – Event Simulation: A First Course, Pearson Education, 2006.

2. Averill M. Law: Simulation Modeling and Analysis, 4 th Edition, Tata McGraw-Hill, 2007

3**.** System Modelling and Response by Ernest O. Doebelin

4. Averill M Law, "Simulation Modeling and Analysis",McGraw-Hill Inc,2007 Geoffrey Gorden, "System Simulation",Prentice Hall of India,1992.

223AGE009	Principles of Automation	CATEGORY	L	T	P	CREDIT
		CREDIT COURSE	3	0	0	0

Preamble:

This course deals in detail with the various aspects of automation such as sensors, actuators, controllers, mechanical and electrical elements and their integration for automating new and existing manufacturing and process industries and applications. This course will be beneficial to students in designing automation schemes for industries and to design automated systems

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

CO 1	Explain the fundamentals of sensor systems and to choose a suitable sensor system for the given application based on the evaluation of the constraints.
CO 2	Explain the fundamentals of signal conditions and to design a suitable signal conditioning scheme for given application.
CO 3	Describe the characteristics of various actuator systems and to decide the right type of actuator for the given application.
CO 4	Describe the importance of an industrial robot and fundamentals of numerical control in automation.
CO 5	Explain the fundamentals of controllers used in industrial automation and to construct simple automation schemes by ladder logic programs.

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	2		2	2	2		
CO 2	2		2	2	2		
CO 3	2		2	2	2		
CO 4	2		2	2	2		
CO 5	2		2				

Assessment Pattern

Bloom's Category	End Semester Examination
Understand	70 %
Apply	30 %

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern: 40 marks
Course based task (Project/Assignments/Simulations/Case studies): 15 marks
Seminar/Quiz: 15 marks
Test paper, 1 no.: 10 marks
Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern:**60 marks**
The examination will be conducted by the respective College. The examination will be for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any five. Each question can carry 12 marks.

Model Question Paper
223AGE009 Principles of Automation

Time 2.5 Hrs Marks 60

Answer any five questions Each carries 12 marks

1. (a) Differentiate the static and dynamic characteristics of a temperature sensor and explain how it affects the selection of a suitable temperature sensor. (6 marks)
   (b) Explain the working of a strain-gauge. (6marks)

2. (a) Explain why anti-aliasing filters are used in analog to digital converters. (3 marks)
   (b) Design a first order low pass filter with a cutoff frequency of 2 kHz. (9 marks)

3. (a) What are the factors to consider while deciding choosing between hydraulic, pneumatic or electrical actuation systems for an automation scheme? (4 marks)
   (b) Explain the working of a three-way pressure reducing valve. (4 marks)
   (c) Explain the working of solenoids. In what applications would you use a Solenoid valve. (4 marks)

4. (a) Explain the principle of the Touch sensor and also mention how they are used in robots. (5 marks)
   (b) Explain the basic terminologies in robotic system and also explain the components of robotic system. (7 marks)

5. (a)With neat schematic explain the architecture of the PLC. (6 marks)
   (b) Explain the use of an up-down counter in PLC with a suitable example. (6 marks)

6. (a) Write short note on SCADA. What is difference PLC and SCADA? (3 marks)

   (b)Construct a ladder logic for controlling a process tank as per the logic given below;
   i.The tank should be filled by a valve V1 when low level float switch L1 is ON and an external input S1 is received.

ii. V1 should be closed when the liquid level reaches a high-level float switch L2.
iii. An agitator motor should be turned on after a delay of 5sec after L2 is triggered.
iv. After agitating for 30mins, contents of the tank should be emptied by opening another valve V2.
v. The temperature should be maintained at 70℃ using a thermostat T1 and Heater H
(9 marks)

7. (a) Explain the levels of Automation. (6 marks)
(b) Explain the working of Flow sensor (6 marks)

Syllabus and Course Plan

No	Topics	No. of Lectures
1	**Introduction to Industrial Automation**	
1.1	Basic Elements of an Automated System, Levels of Automation	2
1.2	Hardware components for Automation: Sensors, classification, Static and dynamic behaviour of sensors.	2
1.3	Basic working principle of different sensors: Proximity sensors, Temperature sensors, flow sensors, Pressure sensors, Force sensors. Position sensors	4
2	**Signal conditioning**	
2.1	Need for signal conditioning, Types of signal conditioning.	2
2.2	Signal conditioning using operational amplifier-Amplifier (Inverting and Non-inverting) and Filter circuits (Basic concepts). Design of first order low pass filter.	2
2.3	Signal conditioning for data acquisition systems, anti-aliasing filters, Analog–Digital Conversions, Analog-to-Digital Converters (ADC)- Steps in analog-to-digital conversion, Successive Approximation Method, Digital-to-Analog Converters (DAC)- Steps in digital to analog conversion, Zero-order and first order data hold circuits	4
3	**Actuators**	
3.1	Types of actuators- mechanical, electrical, pneumatic and hydraulic actuators. (Basic working principle)	2
3.2	Mechanical systems for motion conversion, transmission systems	3
3.3	Solenoids, Electric and stepper motors control.	3
4	**Robotics and Automated Manufacturing Systems**	
4.1	Robot Anatomy and Related Attributes: Joints and Links, Common Robot Configurations, Joint Drive Systems, Sensors in Robotics (Basic concepts)	3
4.2	Robot Control Systems, Applications of Industrial Robots- Material handling	4
4.3	Fundamentals of Numerical control (NC) Technology	1
5	**Discrete Control and Programmable Logic Controllers**	

5.1	Discrete Process Control: Logic and Sequence control	2
5.2	Ladder Logic Diagrams, Programmable Logic Controllers: Components of the PLC, PLC Operating Cycle, Programming the PLC (Basic concepts only)	4
5.3	Introduction to Distributed control system (DCS) and Supervisory Control and Data Acquisition Systems (SCADA)	2

Reference Books

1. Mikell Groover, Automation, Production Systems, and Computer-Integrated Manufacturing, 5th Edition, Pearson, 2019.
2. Yoram Koren, “Computer Control of Manufacturing Systems”,TataMcGraw Hill Edition2005.
3. S. R. Deb; Sankha Deb. Robotics Technology and Flexible Automation, Second Edition McGraw-Hill Education: New York, 2010.
4. W. Bolton, “Mechatronics: Electronic Control Systems in Mechanical and Electrical Engineering” - PrenticeHall- 2013 - 5th Edition.
5. Doebelin, E.O. and Manic, D.N., “Measurement Systems: Applications and Design”, 7th Edition, McGraw Hill, 2019.
6. Krishna Kant, Computer Based Industrial Control-, EEE-PHI,2nd edition,2010.
7. Nathan Ida, Sensors, Actuators, and Their Interfaces- A multidisciplinary introduction, 2nd Edition, IET Digital Library, 2020.
8. Salivahanan, S., and VS Kanchana Bhaaskaran. Linear integrated circuits. McGraw-Hill Education, 2nd edition, 2014.
9. Petruzella, Frank D. Programmable logic controllers. Tata McGraw-Hill Education, 2005
10. Chapman and Hall, “Standard Handbook of Industrial Automation”, Onsidine DM C & Onsidine GDC”, NJ, 2006

223AGE002	FORENSIC ENGINEERING	CATEGORY	L	T	P	CREDIT
		Audit Course	3	0	0	-

Preamble: This course explores various aspects of Forensic Engineering and different methods ,tools and procedures used by Engineers to investigate and analyze . The students will learn to develop their awareness in Forensic Engineering .

Pre-requisite: Nil

Course Outcomes:

After the completion of the course the student will be able to

CO 1	Identify the fundamental aspects of forensic Engineering
CO 2	Apply forensic Engineering in Practical work flow and Investigation
CO 3	Apply methods and analysis in Forensic Investigation
CO 4	Develop practical strategies and standards of Investigation
CO 5	Create an awareness in criminal cases and create Engineering expertise in court room on forensic Engineering

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO7
CO 1	2	2	3	3	3	3	
CO 2	2	2	3	3	3	3	1
CO 3	3	3	3	3	3	3	1
CO 4	3	3	3	3	3	3	1
CO 5	3	3	3	3	3	3	

Assessment Pattern

Bloom's Category	Continuous Internal Evaluation	End Semester Examination
Apply	40 %	60 %
Analyse	40 %	40 %
Evaluate	20 %	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation: 40 marks

Course based task :15marks

Seminar/Quizz :15marks

Test paper :10 marks

Test paper shall include minimum 80% of the syllabus.

End Semester Examination: 60 marks

The examination will be conducted by the respective College. The examination will be for 150 minutes and will contain 7 questions, with minimum one question from each module of which student should answer any five. Each question can carry 12 marks.

Model Question paper

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

THIRD SEMESTER M. TECH DEGREE EXAMINATION

Course Code: 223AG002

Course Name: FORENSIC ENGINEERING

Max. Marks: 60 **Duration: 2.5 Hours**

PART A

Answer any 5 questions, each question carries 12 marks. **Marks**

1.	(a)	What are the uses of forensic engineering in legal laws ?	(7)
	(b)	Discuss the professional responsibility of a forensic Engineer .	(5)
2.	(a)	What are the steps in preliminary on site Investigation ?	(7)
	(b)	With suitable examples, explain photo cataloguing?	(5)
3.	(a)	Discuss STEP method .	(7)
	(b)	Explain root cause Analysis	(5)
4.	(a)	Detail about EDAX Method.	(7)
	(b)	Enlist the uses of NDT in forensic Analysis with example	(5)
5.	(a)	Differentiate NFPA & FMV Standards	(7)
	(b)	Briefly discuss the term Email Phishing ?	(5)
6.		Define the responsibility and duty of a forensic expert in the court.	(12)
7.		Explain Forensic Engineering workflow with examples	(12)

Syllabus and Course Plan

Module No	Topic	No. of Lectures (Hours)
1	**Module 01: Introduction to Forensic Engineering (6 Hours)**	
1.1	Forensic Engineering-Definition, Investigation Pyramid, Eyewitness Information, Role in Legal System	2
1.2	Scientific Method-Applying scientific methods in Forensic Engineering-Engineer as expert Witness-Scientific methods and legal system	2
1.3	Qualification of Forensic Engineer-Technical- Knowledge- Oral-written-Communication- other skills-Personality Characteristics	1
1.4	Ethics and professional responsibilities.	1
2	**Module 02: Forensic Engineering Workflow and Investigation Methods (6 Hours)**	
2.1	Forensic Engineering Workflow-Team &planning-preliminary onsite investigation. Sampling-selection of sample-collection- packing-sealing of samples.	2
2.2	Source and type of evidence - Paper documentation- digital documentation-electronic data. Physical Evidence-Collection of photograph-cataloguing -Recognizing the Evidence-organizing-Evidence Analysis -Reporting	2
2.3	Investigation Methods- Cause and Causal mechanism analysis-Time and event sequence-STEP method. Human Factors, Human errors - Analysis of Operative Instruction and working Procedures	2
3	**Module 03: Physical Product Failure & Analytical Methods (6 Hours)**	
3.1	Introduction to typical Forensic Engineering Tool box-NDT, Crack detection and human eye -Hardness testing- and Destructive testing Methods with case studies	2
3.2	Indirect stress strain Analysis-Brittle lacquer technique, Contact Radiography-Metallography-EDAX method	1
3.3	Forensic Optical Microscopy-Examination- Magnification-USB Microscopy -Wifi Enabled microscopy -Reflected microscopy	2
3.4	Novel Tools and System -Contour Method-Flash Thermography-Thermographic signal reconstruction (TSR)-Electromagnetically induced acoustic Emission (EMAE)-Pulsed Eddy Current (PEA)-Theory only	1
4	**Module 04: Cyber Forensic , Civil ,Electrical Accidents & Standards (6 Hours)**	
4.1	Basics of Digital & Cyber forensics: Technical concepts; labs and tools; collecting evidence Operating System Forensic basics with - Windows, Linux -Mobile Forensic-Anti forensics-Malware- Web attack forensics with Email Crimes-Cyber Laws	3
4.2	Different types of Forensic accident investigations- Civil Engineering-Structural- Road accidents -Fire accidents - Water related accidents-Electrical accidents and Investigation methods	2
4.3	Protocol for forensic Investigations-Standard guides-scope significance -use -procedures- reports. Standards – ASTM standards -FMV Standards -SAE Standards -Relevant Standards -NFPA Standards -International Standards	1

5	**Module 05: Engineer in the Court room& Criminal Cases (6 Hours)**	
5.1	Role of an Engineering Expert-Report-pre trial meetings-Alternative dispute resolution-Single joint expert. Engineer in the court room	2
5.2	Criminal Cases-Introduction-Counterfeit coins-fraudulent road accidents-Fraudulent Insurance claims.	2
5.3	Cyber Crimes and Cases- SIM Swapping -ATM Cloning-Microsoft Internal Spam- Intellectual property cases.	2

Reference Books

1. Colin R Gagg, *Forensic EngineeringThe Art &Craft of a failure detective* , Taylor & Francis Publishing, 2020
2. Luca Fiorentini ,Luca Marmo *Principles of Forensic Engineering Applied to Industrial Accidents* , Wiley, 2019
3. Harold Franck, Darren Franck , *Forensic Engineering Fundamentals* ,Taylor & Francis publishing 2013
4. Randall K Noon , *Forensic Engineering Investigation*, CRC press limited , 2001
5. Stephen E Petty , *Forensic Engineering: Damage assessment for residential and commercial structures* CRC press 2nd edition , 2017
6. Joshua B Kardon , *Guideliness for forensic Engineering practice* , ASCE, 2012
7. Richard W. Mclay and Robert N. Anderson, *Engineering standards for forensic Applications* , Academic Press; 1st edition 2018
8. Max M Houck ,*Forensic Engineering (Advanced forensic Science),* Academic press 1st edition 2017
9. Niranjan Reddy - Practical Cyber Forensics. *An Incident-based Approach to Forensic Investigations-Apress (2019)*
10. Peter Rhys Lewis, Ken Reynolds, Colin Gagg - *Forensic Materials Engineering Case Studies-CRC Press (2003) (1)*

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

CET campus, Thiruvananthapuram - 695 016

Ph: 0471 2598122; Fax: 2598522 www.ktu.edu.in Email: university@ktu.edu.in

File Ref.No.KTU/ASST6(ACADEMIC)/6559/2023

APJ Abdul Kalam Technological University
Thiruvananthapuram

Abstract

Academic - M. Tech 2022 Regulations - Dissertation Phase I and Phase II (Track I) Syllabus approval and Mode of Evaluation for Internship - orders issued - reg

---

ACADEMIC SECTION

U.O.No. 3070/2023/KTU Thiruvananthapuram, Dated: 25.11.2023

Read:-1. Minutes of the meeting of the Board of Studies (PG) held on 15/11/2023.

ORDER

Vide paper read as above, the Board of Studies (PG) has finalized the syllabus for the M. Tech Dissertation Phase I in Semester III and Dissertation Phase II in Semester IV of Track I, 2022 Regulations.

The meeting of the Board of Studies (PG) pointed out the difficulty in conducting the end-semester viva voice for the Internship in Semester III with external members and recommended to conduct the end-semester evaluation of the internship internally. This internal evaluation would be carried out by a committee comprising the Faculty Supervisor, PG Program Coordinator, and one faculty member from a sister department.

Sanction is accorded to implement the recommendations of the Board of Studies (PG) for approving the syllabus for the M. Tech Dissertation Phase I (Semester III) and Dissertation Phase II (Semester IV) of Track I, attached as annexure and the Mode of evaluation for Internship in Semester III, 2022 Regulations, subject to ratification by the Academic Council.

This order is issued under Sub Section (5) of Section 14 of the APJ Abdul Kalam Technological University Act, 2015 (17 of 2015).

Orders are issued accordingly.

Sd/-

Dr Saji Gopinath *
Vice Chancellor

Copy to:-

1.KTU Affiliated Colleges
2.JD(IT)for necessary action
3.PA to Dean Academic
4.PA to Registrar

* This is a computer system (Digital File) generated letter. Hence there is no need for a physical signature.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
223PXX100	DISSERTATION PHASE I	Project Work	0	0	17	11

COURSE OBJECTIVES:

Dissertation is aimed to bridge the gap between theoretical knowledge and practical application, fostering a well-rounded skill set that prepares students for success in their future engineering careers. Engineering projects often simulate real-world engineering scenarios. This exposure allows students to become familiar with industry practices, standards, and expectations and preparing them for the challenges they might face in their future careers. Depending on the nature of the project, students may acquire practical skills related to specific tools, software, or equipment. This hands-on experience can be highly beneficial when transitioning to a professional engineering role.

Dissertation Phase I can help to identify the problem based on the area of interest through proper literature survey and to foster innovation in design of products, processes or systems based on the identified problem. perform feasibility study by creative thinking and requirement analysis in finding viable solutions to engineering problems

All categories of students in track 1 are to carry out the dissertation in the Institute they are studying or in any CSIR/Industrial/ R&D organization/any other reputed institute which have facilities for dissertation work in the area proposed.

Course Outcomes:

After the completion of the course the student will be able to

CO 1	Identify and define a relevant and significant problem or challenge in the relevant field
CO2	Formulate research methodologies for the innovative and creative solutions
CO 3	Plan and execute tasks utilizing available resources within timelines, following ethical professional and financial norms
CO 4	Organize and communicate technical and scientific findings effectively in written reports, oral presentation, and visual aids

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO7
CO 1	3		3	2	2	3	2
CO 2	3		3	3	3	2	
CO 3	3		2		3	3	2
CO 4		3	3	2			2

Continuous Internal Assessment (CIA) Total Marks: 100
The evaluation committee comprises

1- Project Coordinator(s)
2- A Senior faculty member
3- Supervisor of the student

Pattern:

Zeroth evaluation by the Evaluation Committee	-
Interim evaluation by the Evaluation Committee	20 marks
Final evaluation by the Evaluation Committee	40 marks
Project Phase - I Report (By Evaluation Committee)	20 marks
Project progress evaluation by supervisor	20 marks

The Plagiarism level in the project report shall be less than 25%.

Interim Review

Literature Survey (CO1- 5 marks)
Comprehension and Problem Identification (CO2-5 marks)
Objective Identification (CO2-5 marks)
Document Preparation and Presentation (CO4-5 marks)

Final Review

Literature Survey (CO1-10 marks)
Project Design (CO2-10 marks)
Execution of tasks by utilizing available resources within timelines (CO3 – 10 marks)
Presentation and document preparation (CO4-10 marks)

Evaluation by the supervisor

The guide/supervisor shall monitor the progress being carried out by the student on a regular basis. In case it is found that progress is unsatisfactory it shall be reported to the Department Evaluation Committee for necessary action.

Student's Diary/ Log book: The main purpose of writing diary/log book is to cultivate the habit of documenting and to encourage the students to search for details. The activity diary shall be signed after every week by the supervisor.

The minimum attendance for completing the course is 75%. The pass minimum for the course is 50% for CIA.

SYLLABUS:

DETAILS	HOURS
` 1. Literature study/survey of published literature on the assigned topic 2. Formulation of objectives 3. Formulation of hypothesis/ design/ methodology 4. Formulation of work plan and task allocation. 5. Design documentation 6. Preliminary analysis/Modelling/Simulation/Experiment/Design/Feasibility study 7. Preparation of Phase 1 report	150

Dissertation outside the Institute: For doing dissertation outside the Institution, the following conditions are to be met:

i. They have completed successfully the course work prescribed in the approved curriculum up to the second semester.
ii. The student has to get prior approval from the DLAC and CLAC.
iii. Facilities required for doing the dissertation shall be available in the Organization/Industry (A certificate stating the facilities available in the proposed organization and the time period for which the facilities shall be made available to the student, issued by a competent authority from the Organization/Industry shall be submitted by the student along with the application).
iv. They should have an external as well as an internal supervisor. The internal supervisor should belong to the parent institution and the external supervisor should be Scientists or Engineers from the Institution/Industry/ R&D organization with which the student is associated for doing the dissertation work. The external supervisor shall be with a minimum post graduate degree in the related area.
v. The student has to furnish his /her monthly progress as well as attendance report signed by the external supervisor and submit the same to the concerned Internal supervisor.
vi. The external supervisor is to be preferably present during all the stages of evaluation of the dissertation.

Internship leading to Dissertation: The M. Tech students who after completion of 6 to 8 weeks internship at some reputed organizations are allowed to continue their work as dissertation for the third and fourth semester after getting approval from the CLAC. Such students shall make a brief presentation regarding the work they propose to carry out before the DLAC for a detailed scrutiny and to resolve its suitability for accepting it as an M.Tech dissertation. These students will be continuing as regular students of the Institute in third semester for carrying out all academic requirements as per the curriculum/regulation. However, they will be permitted to complete their dissertation in the Industry/Organization (where they have successfully completed their internship) during fourth semester. They should have an external as well as an internal supervisor. The internal supervisor should belong to the parent institution and the external supervisor should be Scientists or Engineers from the external organization with which the student is associated for doing the dissertation work. The external supervisor shall be with a minimum post graduate degree in the related area. The student has to furnish his /her monthly progress as well as attendance report signed by the external guide and submit the same to the concerned internal guide. The external guide is to be preferably present during all the stages of evaluation of the dissertation.

Dissertation as part of Employment: Students may be permitted to discontinue the programme and take up a job provided they have completed all the courses till second semester (FE status students are not permitted) prescribed in the approved curriculum. The dissertation work can be done during a later period either in the organization where they work if it has R&D facility, or in the Institute. Such students should submit application with details (copy of employment offer, plan of completion of their project etc.) to the Dean (PG) through HoD. The application shall be vetted by CLAC before granting the approval. When the students are planning to do the dissertation work in the organization with R&D facility where they are employed, they shall submit a separate application having following details:

i. Name of R&D Organization/Industry
ii. Name and designation of an external supervisor from the proposed Organization/Industry (Scientists or Engineers with a minimum post graduate degree in the related area) and his/her profile with consent
iii. Name and designation of a faculty member of the Institute as internal supervisor with his/her consent
iv. Letter from the competent authority from the Organization/Industry granting permission to do the dissertation
v. Details of the proposed work
vi. Work plan of completion of project

DLAC will scrutinize the proposal and forward to CLAC for approval. When students are doing dissertation work along with the job in the organization (with R & D facility) where they are employed, the dissertation work shall be completed in four semesters normally (two semesters of dissertation work along with the job may be considered as equivalent to one semester of dissertation work at the Institute). Extensions may be granted based on requests from the student and recommendation of the supervisors such that he/she will complete the M. Tech programme within four years from the date of admission as per the regulation. Method of assessment and grading of the dissertation will be the same as in the case of regular students. The course work in the 3rd semester for such students are to be completed as per the curriculum requirements (i) MOOC can be completed as per the norms mentioned earlier (ii) Audit course are to be carried out either in their parent Institution or by self-learning. However, for self-learning students, all assessments shall be carried out in their parent institution as in the case of regular students.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
224PXX100	**DISSERTATION PHASE II**	**Project Work**	**0**	**0**	**24**	**16**

All categories of students in track 1 are to carry out the DISSERTATION PHASE II in the institute they are studying or in any Industrial/ R&D organization/any other reputed institute which have facilities for dissertation work in the area proposed. DISSERTATION PHASE II shall not compulsorily continuation of DISSERTATION PHASE I. The student has to publish a research article in a conference or a reputed journal before appearing for the end-semester examination. The eligibility criteria for registering to the end semester examination are attendance in the course and no pending disciplinary action. The minimum attendance for appearing for the end semester examination is 75%. Students who do not meet these eligibility criteria are ineligible (identified by FE grade) to appear for the ESE. Students, who have completed a course but could not appear for the end semester examination, shall be awarded 'AB' Grade, provided they meet other eligibility criteria The pass minimum for the course is 45% for ESE and 50% for (CIA and ESE) put together.

Continuous Internal Assessment (CIA) Total Marks: 100

The evaluation committee comprises

1- Project Coordinator(s)
2- A Senior faculty member
3- Supervisor of the student

Pattern (CIA)

Zeroth evaluation by the Evaluation Committee	-
Interim evaluation by the Evaluation Committee	30 marks
Final evaluation by the Evaluation Committee	50 marks
Project progress evaluation by supervisor	20 marks

Evaluation by the supervisor

The guide/supervisor shall monitor the progress being carried out by the student on a regular basis. In case it is found that progress is unsatisfactory it shall be reported to the Department Evaluation Committee for necessary action.

Student's Diary/ Log book: The main purpose of writing diary/log book is to cultivate the habit of documenting and to encourage the students to search for details. The activity diary shall be signed after every week by the supervisor.

End Semester Evaluation (ESE) Total Marks: 100

The evaluation committee comprises

1- Project Coordinator(s)
2- An external expert (from Industry or research/academic institute)
3- Supervisor of the student

Pattern (ESE)

1. **Innovation and Originality (10 marks):**
Assessment of the uniqueness and innovation demonstrated in the project work.
Original contributions, if any, to the field or problem area.
2. **Implementation and Execution (20 marks):**
Evaluation of the actual implementation or execution of the project, including:
    - Quality of work done
    - Demonstrated skills and techniques applied
    - Adherence to project timelines and milestones
3. **Project Documentation (25 marks):**
Comprehensive project report evaluation including:
    - Introduction and problem statement
    - Literature review
    - Methodology and approach
    - Results and analysis
    - Conclusion and recommendations
    - References and citations
    - Details of the publications
    - Plagiarism certificate

The Plagiarism level in the project report shall be less than 25%.

4. **Presentation and Defence (40 marks):**
Oral presentation of the project to a panel of examiners, including:
    - Clarity and effectiveness of the presentation
    - Ability to explain the project objectives, methodologies, and findings
    - Handling questions and providing satisfactory answers during the defence

5. **Publication of the work either in a conference or in a journal (5 marks)**

SYLLABUS:

DETAILS	HOURS
` 1. Literature study/survey of published literature on the assigned topic 2. Topic Selection and Proposal 3. Formulation of objectives 4. Research and Planning 5. Formulation of work plan and task allocation. 6. Execution 7. Documentation and Reporting 8. Project Showcase reflecting on the project experience and lessons learned	200

Dissertation outside the Institute: For doing dissertation outside the Institution, the following conditions are to be met:

i. They have completed successfully the course work prescribed in the approved curriculum up to the second semester.
ii. The student has to get prior approval from the DLAC and CLAC.
iii. Facilities required for doing the dissertation shall be available in the Organization/Industry (A certificate stating the facilities available in the proposed organization and the time period for which the facilities shall be made available to the student, issued by a competent authority from the Organization/Industry shall be submitted by the student along with the application).
iv. They should have an external as well as an internal supervisor. The internal supervisor should belong to the parent institution and the external supervisor should be Scientists or Engineers from the Institution/Industry/ R&D organization with which the student is associated for doing the dissertation work. The external supervisor shall be with a minimum post graduate degree in the related area.
v. The student has to furnish his /her monthly progress as well as attendance report signed by the external supervisor and submit the same to the concerned internal supervisor.
vi. The external supervisor is to be preferably present during all the stages of evaluation of the dissertation.