ADT302	CONCEPTS IN BIG DATA ANALYTICS	Category	L	T	P	Credits	Year of Introduction
		PCC	3	1	0	4	2020

Preamble: This course helps the learner to understand the basic concepts of big data analytics. This course covers on big data technologies used for storage, analysis and manipulation of data. The student will learn about fundamentals of Hadoop, MapReduce, Pig, Hive, R and have hand on training on the same It also help to develop projects and apply existing data analytics tools to gain comprehensive knowledge on Data analytics. It enables the learners to perform data analysis on a real-world scenario using appropriate tools.

Prerequisite : Basic knowledge in programming

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

CO#	Course Outcomes
CO1	Outline the basic big data concept. **(Cognitive KnowledgeLevel: Understand)**
CO2	Categorize and summarize the processing in Big Data and its importance. **(Cognitive Knowledge Level:Understand)**
CO3	Simulate various Big data technologies like Hadoop MapReduce, Pig, Hive, Hbase.(**CognitiveKnowledge Level: Apply)**
CO4	Determine tools and techniques to analyze Big Data **(CognitiveKnowledge Level: Apply)**
CO5	Resolve problems associated with big data with the features of R programming **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

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

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

Assessment Pattern

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

Mark Distribution

Total marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	4

Continuous Internal Evaluation Pattern:

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

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. 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 willbe two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly completed module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any5.

End Semester Examination Pattern:

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

a maximum 2 sub-divisions and carries 14 marks.

SYLLABUS

Module – 1 (Introduction to Big Data)

Introduction to Big data, Conventional Data vs Big data, Big data architecture, Big data platforms, Nature of data, Analytic processes and tools, 5 V's of Big data, Big data analytical method, Intelligent data analysis, Big data analytics life cycle.

Module - 2 (Introduction to Stream Computing)

Introduction to stream concepts – Streaming data architecture, Stream data model, Sampling techniques for efficient stream processing, Filtering streams – Bloom filter, Count distinct problem – Flajolet martin algorithm, Estimating moments, Counting oneness in a window – DGIM Algorithm

Module - 3 (Hadoop Distributed File System)

History of Hadoop, Hadoop Ecosystem, Core Components, HDFS- Architecture, Using HDFS Files, HDFS Design, Blocks, Namenodes and Data nodes, Basic File system Operations, Hadoop Specific File Types, Anatomy of a file read, Anatomy of a file write. Data Processing with MapReduce: Execution Pipeline, Runtime Coordination and Task Management in MapReduce, Designing MapReduce implementations: Using MapReduce as a framework for parallel processing, Example-Road Enrichment.

Module - 4 (Pig, Hive, HBase)

Pig : Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators. Hive : Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions. Hbase : HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.

Module - 5 (Introduction to R programming)

Introduction to R – Overview of modern data analytic tools, Introduction to R, R Graphical User Interfaces - Features of R Language, Vectors, Filtering, Creating Matrices , Applying Functions to Matrix Rows and Columns, Lists , Creating List , General List Operations, Data Frames , Creating Data Frames , Matrix like Operations in Frames , Applying Functions to Data Frames ,Reading and Writing Files.

Text Book

1. Tom White " Hadoop: The Definitive Guide" Third Edit on, O'reily Media, 2012.
2. Michael Minelli, Michelle Chambers and AmbigaDhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013.
3. Boris Lublinsky, Kevin T. Smith, Alexey Yakubovich ,Professional Hadoop Solutions.

4. Norman Matloff , “The Art of R Programming: A Tour of Statistical Software Design”, NoStarch Press.

References Books

1. Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, 1st Edition, Wiley and SAS Business Series,2012.
2. Jure Leskovec, Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2014.
3. Seema Acharya, Subhasni Chellappan, “Big Data And Analytics”, Wiley Publications.
4. BIG DATA, Black Book TM, DreamTech Press, 2016 Edition.
5. Nathan Marz and James Warren, “BIG DATA- Principles and Best Practices of Scalable Real-time Systems”.
6. Jason Rutherglen, Dean Wampler, Edward Capriolo, Programming Hive, O'Reilly.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Explain the features of the integrated IT solution for Big data management.
2. Define the term “Big data”. How do 5 V’s help to decide whether a given data source contributes to big data.
3. Identify the differences between data analysis and data reporting.

Course Outcome 2 (CO2):

1. Some websites check availability of username by searching millions of usernames registered with it. Identify one effective method to filter data as in this type of scenario.
2. Discuss the issues in stream processing.
3. An array consists of some elements A=8,10,… and the size of array is set to 10. Check whether 96, 21 lies in the array or not. [Hash functions: 3x+3 mod6, 3x+7mod8, 2x+9 mod2, 2x+3mod5]

Course Outcome 3 (CO3):

1. Explain the components of Hadoop.
2. Illustrate map reduce job execution flow.
3. Explain HBase client ecosystem.

Course Outcome 4 (CO4):

1. Explain two execution types or modes in PIG.
2. Summarize any three relational operations in Pig Latin with examples
3. Illustrate managed tables and external tables in HIVE.

Course Outcome 5 (CO5):

1. Illustrate any three R functions used in data analytics.
2. Explain the different categories of attributes and data types in R.
3. Write a short note about how the different types of files can be read and write in R.

Model Question Paper ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

QP CODE: **PAGES : 3**

Reg No: _____Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: ADT302

Course Name: Concepts In Big Data Analytics

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

PART A

Answer All Questions. Each Question Carries 3 Marks

1. How are big data and hadoop related to each other?
2. What are the 5 Vs of Big Data
3. Explain the features and column families of HBase.
4. Compare the specific file types of HDFS.
5. How does Map Reduce Framework provide support for application development?
6. Describe the Map Reduce job implementation in the case of Road Enrichment example.
7. Describe Filtering Streams.
8. Explain about the partitioned and managed tables in Hive.
9. Identify the ways in which a pig program can be executed.
10. Discuss the general list operations in R with example. **(10x3=30)**

Part B

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

11. Illustrate Big Data Architecture. **(10)**

(a)

(b) Compare conventional Data and Big Data **(4)**

OR

12. (a) Explain the life cycle of big data analytics in detail. **(10)**

(b) Compare the types of Big Data with examples. **(4)**

13. (a) Suppose we have a window of length N (say N=24) on a binary system, we want at all times to be able to answer a query of the form " How many 1's are there in the last K bits?" for K<=N. Suggest an algorithm to solve this issue with detailed explanation. Find the total number of ones, when 0111 enters into the given stream ….10101100010111010110010110…. (Assume, the new **(8)**

bit enters from the right side and time stamp of first new bit is 100)

(b) Write the advantages and disadvantages of Data Stream. **(8)**

OR

14. (a) Illustrate the working of Bloom filter with examples for **(10)**
i) Inserting an element
ii) Searching an element.

(b) Why is conventional data processing insufficient for stream processing? **(4)**

15. (a) Explain the data model and architecture of HBase. **(10)**

(b) Discuss on the general guidelines for HBase Schema Design. **(4)**

OR

16. (a) Explain the anatomy of MapReduce Job run using classic MapReduce **(6)**

(b) Explain the types of Schedulers available in YARN. **(8)**

17. (a) Explain the main components of Hadoop Pig framework. **(4)**

(b) Write the syntax to create a table and partition in Hive. **(10)**

OR

18. (a) Describe about Data Types and File Formats in Hive. **(8)**

(b) Write about Pig Latin Structure and functions **(6)**

19. (a) Explain in detail about the Matrix handling in R. **(8)**

(b) List and explain four R functions used in descriptive statistics. **(6)**

OR

20. (a) Discuss the data visualization for multiple variables in R **(8)**

(b) Describe the R functions used for cleaning dirty data **(6)**

Teaching Plan

No	Contents	No of Lecture Hrs (45)
	Module – 1(Introduction to Big Data) (9 hrs)	
1.1	Introduction to Big data, Conventional Data vs Big data	1
1.2	Big data architecture	1
1.3	Big data platforms	1
1.4	Nature of data,	1
1.5	Analytic processes and tools.	1
1.6	5 V's of Big data	1
1.7	Big data analytical method	1
1.8	Intelligent data analysis	1
1.9	Big data analytics life cycle	1
	Module – 2 (Introduction to Stream Computing) (8 hrs)	
2.1	Introduction to stream concepts	1
2.2	Streaming data architecture.	1
2.3	Stream data model	1
2.4	Sampling techniques for efficient stream processing	1
2.5	Filtering streams – Bloom filter	1
2.6	Count distinct problem - Flajolet martin algorithm	1
2.7	Estimating moments	1
2.8	Counting oneness in a window – DGIM algorithm	1
	Module - 3 (Hadoop Distributed File System) (13 hrs)	
3.1	History of Hadoop	1

3.2	Hadoop Ecosystem and Core Components	1
3.3	HDFS Architecture	1
3.4	Using HDFS Files ,HDFS Design	1
3.5	Blocks, Namenodes and Data nodes	1
3.6	Basic File system Operations	1
3.7	Hadoop Specific File Types	1
3.8	Anatomy of a file read	1
3.9	Anatomy of a file write	1
3.10	Execution pipeline	1
3.11	Runtime Coordination and Task Management in MapReduce	1
3.12	Using MapReduce as a framework for parallel processing	1
3.13	Road Enrichment Example	1
	Module - 4 (Pig, Hive, Hbase) (7 hrs)	
4.1	Pig : Introduction to PIG, Execution Modes of Pig	1
4.2	Comparison of Pig with Databases, Grunt.	1
4.3	Pig Latin, User Defined Functions, Data Processing operators	1
4.4	Hive : Hive Shell, Hive Services	1
4.5	Hive Metastore, Comparison with Traditional Databases.	1
4.6	HiveQL, Tables, Querying Data and User Defined Functions.	1
4.7	Hbase : HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.	1
	Module - 5 (Introduction to R programming) (8 hrs)	
5.1	Introduction to R – Overview of modern data analytic tools, Introduction to R, R Graphical User Interfaces	1
5.2	Features of R Language, Vectors	1
5.3	Filtering and Creating Matrices	1

5.4	Applying Functions to Matrix Rows and Columns	1
5.5	Creating List and General List Operations	1
5.6	Examining Multiple Variable	1
5.7	Creating Data Frames and Matrix like Operations in Frames	1
5.8	Applying Functions to Data Frames and Reading and Writing Files	1

AIT304	ROBOTICS AND INTELLIGENT SYSTEM	Category	L	T	P	Credit	Year of Introduction
		PCC	3	1	0	4	2022

Preamble: This course enables the learners to understand the fundamental concepts and algorithms in Robotics and Intelligent systems. The course covers the standard hardware and kinematic concepts for robot design. Standard algorithms for localization, mapping, path planning, navigation and obstacle avoidance, to incorporate intelligence in robots are included in the course. This course helps the students to design robots with intelligence in a real world environment.

Prerequisite: Basic understanding of probability theory, linear algebra, machine learning, artificial intelligence

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

CO1	Understand the concepts of manipulator and mobile robotics. **(Cognitive Knowledge Level: Understand)**
CO2	Choose the suitable sensors, actuators and control for robot design. **(Cognitive Knowledge Level: Apply)**
CO3	Developing kinematic model of mobile robot and understand robotic vision intelligence. **(Cognitive Knowledge Level: Apply)**
CO4	Discover the localization and mapping methods in robotics. **(Cognitive Knowledge Level: Apply)**
CO5	Plan the path and navigation of robot by applying artificial intelligence algorithm. **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

	PO 1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO1 0	PO11	PO1 2
CO1	✓				✓	✓						✓
CO2	✓	✓			✓							✓
CO3	✓	✓		✓	✓	✓						✓
CO4	✓			✓	✓	✓						✓
CO5	✓			✓	✓	✓						✓

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

Assessment Pattern

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

Mark Distribution

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

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

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

Continuous Assessment Assignment **15 marks**

Internal Examination Pattern

Each of the two internal examinations has to be conducted out of 50 marks. First series test shall be preferably conducted after completing the first half of the syllabus and the second series test shall be preferably conducted after completing remaining part of the syllabus. There will be two

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

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

SYLLABUS

Module – 1 (Introduction to robotics)

Introduction to robotics – Degrees of freedom, Robot types- Manipulators- Anatomy of a robotic manipulator-links, joints, actuators, sensors, controllers. Robot configurations-PPP, RPP, RRP, RRR. Mobile robots- wheeled, legged, aerial robots, underwater robots, surface water robots . Dynamic characteristics- speed of motion, load carrying capacity & speed of response. Introduction to End effectors - mechanical grippers, special tools, Magnetic grippers, Vacuum grippers, adhesive grippers, Active and Passive grippers. Ethics in robotics - 3 laws - applications of robots.

Module - 2(Sensors, Actuators and Control)

Sensor classification- touch, force, proximity, vision sensors. Internal sensors-Position sensors, velocity sensors, acceleration sensors, Force sensors; External sensors-contact type, non contact type; Digital Camera - CCD camera - CMOS camera - Omnidirectional cameras
Sensor characteristics. Actuators - DC Motors - H-Bridge - Pulse Width Modulation - Stepper Motors – Servos, Hydraulic & pneumatic actuators. Control - On-Off Control - PID Control - Velocity Control and Position Control

Module – 3 (Robotic vision & Kinematics)

Robotic Vision: Sensing, Pre-processing, Segmentation, Description, Recognition, Interpretation, Feature extraction -Camera sensor hardware interfacing. Representation of Transformations - Representation of a Pure Translation - - Pure Rotation about an Axis - Combined Transformations - Transformations Relative to the Rotating Frame.

Basic understanding of Differential-Drive Wheeled Mobile Robot, Car-Like Wheeled Mobile Robot. Kinematic model of a differential drive and a steered mobile robot, Degree of freedom and manoeuvrability, Degree of steerability, Degree of mobility - different wheel configurations, holonomic and nonholonomic robots. Omnidirectional Wheeled Mobile Robots.

Module - 4 (Localization and Mapping)

Position and Orientation - Representing robot position. Basics of reactive navigation; Robot Localization, Challenges in localization - An error model for odometric position estimation

Map Representation - Continuous representations - Decomposition strategies - Current challenges in map representation. Probabilistic map-based localization (only Kalman method), Autonomous map building, Simultaneous localization and mapping (SLAM) - Mathematical definition of SLAM - Visual SLAM with a single camera - Graph-based SLAM - Particle filter SLAM - Open challenges in SLAM

Module - 5 (Path Planning and Navigation)

Path Planning- Graph search, deterministic graph search - , breadth first search - depth first search- Dijkstra' s algorithm, A*, D* algorithms, Potential field based path planning. Obstacle avoidance - Bug algorithm - Vector Field Histogram - Dynamic window approaches. Navigation Architectures - Modularity for code reuse and sharing - Control localization - Techniques for decomposition. Alternatives for navigation - Neural networks - Processing the image - Training the neural network for navigation - Convolutional neural network robot control implementation

Text Books

1. R Siegwart, IR Nourbakhsh, D Scaramuzza, Introduction to Autonomous Mobile Robots ,, MIT Press, USA, 2011
2. Thomas Bräunl - Embedded Robotics,Mobile Robot Design and Applications with Embedded Systems-Springer (2006)
3. S.G. Tzafestas - Introduction to Mobile Robot Control-Elsevier (2014)
4. Francis X. Govers - Artificial Intelligence for Robotics-Packt Publishing (2018)
5. Saeed B. Niku - Introduction to Robotics_ Analysis, Control, Applications

Reference Books

1. John J. Craig, Introduction to Robotics, Pearson Education Inc., Asia, 3rd Edition, 2005
2. S. K. Saha, Introduction to Robotics 2e, TATA McGraw Hills Education (2014)
3. Peter Corke - Robotics, Vision and Control_ Fundamental Algorithms in MATLAB® - Springer-Verlag Berlin Heidelberg (2021)

Course Level Assessment Questions

Course Outcome1 (CO1):

1. Categorise the various types of Grippers used in robot manipulators.
2. Differentiate between active and passive grippers.
3. Explain speed of motion and load carrying capacity of a mobile robot.
4. You wish to build a dynamically stable robot with a single wheel only. For each of the four basic wheel types, explain whether or not it may be used for such a robot.

Course Outcome 2(CO2):

1. Categorise the sensors used in robotics
2. Explain any four characteristics of a sensor
3. Illustrate the sensor performance measuring parameters
4. Suggest any two mechanism to realise 360^{o} Camera

Course Outcome 3(CO3):

1. Determine the degrees of mobility, steerability, and maneuverability for each of the following: (a) bicycle; (b) dynamically balanced robot with a single spherical wheel (c) automobile.
2. A frame F was rotated about the y-axis 90^{o}, followed by a rotation about the o-axis of 30^{o}, followed by a translation of 5 units along the n-axis, and finally, a translation of 4 units along the x-axis. Find the total transformation matrix.
3. Explain the camera sensor hardware interfacing.
4. What is an omni directional robot? Explain two configurations to set up an omni directional robot.

Course Outcome 4(CO4): .

1. Explain the challenges of localization
2. How Kalman method can be used in localization of mobile robots
3. What are the Decomposition strategies in map representation
4. How Visual SLAM can be performed with a single camera

Course Outcome 5(CO5):

1. Explain Dijkstra's algorithm with a suitable example.
2. Identify the steps of Generic temporal decomposition of a navigation architecture.
3. What is meant by control decomposition? Explain two types of control decomposition.
4. Why does SLAM work better with wheel odometer data available?

5. In the Floor Finder algorithm, what does the Gaussian blur function does to improve the results?

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AIT304

Course Name: ROBOTICS AND INTELLIGENT SYSTEM

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

PART A

Answer All Questions. Each Question Carries 3 Marks

1. What do you mean by degrees of freedom? How many degrees of freedom are required for a drone to achieve any position in 3D space? And how many more DOF required for achieving any orientation as well.
2. Explain how leg configuration affects the stability of mobile robot.
3. Explain Dynamic range, Linearity and Resolution of a Sensor.
4. Explain the working of a Mechanical accelerometer with a block diagram
5. Differentiate between holonomic and nonholonomic robots.
6. What is the significance of differential drive in mobile robot?
7. How will you represent the position and orientation of a wheeled mobile robot?

8. Identify the 2 mobile robot localization problems.

9. Explain the Bug algorithm for obstacle avoidance. .

10. What is Voronoi diagram method and its advantages? **(10x3=30)**

Part B

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

11. (a) Explain the general features of wheeled, legged and aerial robots. **(9)**

(b) Explain the anatomy of a robotic manipulator with a neat diagram. **(5)**

OR

12. (a) Briefly explain the dynamic characteristics of robots. **(9)**

(b) Assume an object of mass 140 kg is to be lifted up with an acceleration of 10 m/s2. Calculate the gripper force required for the operation, if coefficient of friction between contact surfaces is 0.2, number of fingers in gripper is 2 and acceleration due to gravity is 9.8 m/s2 **(5)**

13. (a) Explain the working of an Optical Encoder. **(5)**

(b) A mobile robot is designed for unidirectional motion with constant velocity. Illustrate the mechanism to make the robot move in forward and reverse direction with variable speed. Support with necessary diagrams **(9)**

OR

14. (a) Compare and contrast the working of CCD and CMOS camera **(9)**

(b) Illustrate the significance of the PID controller with a neat block diagram **(5)**

15. (a) Outline the seven stages of robot vision**.** **(14)**

OR

16. (a) Derive the kinematic model of a differential drive mobile robot. **(7)**

(b) A frame B was rotated about the x-axis 90^o, then it was translated about the current a-axis 3 inches before it was rotated about the z-axis 90^o. Finally, it was translated about the current o–axis 5 inches. **(7)**

(a) Write an equation that describes the motions.

(b) Find the final location of a point p(1,5,4)T attached to the frame relative to the reference frame.

17. (a) Derive error model for odometric position estimation **(8)**

(b) Illustrate the SLAM problem with suitable diagrams **(6)**

OR

18. (a) Compare and Contrast graph based and particle SLAM **(8)**

(b) Describe the concept of mobile robot localization with suitable Block diagrams **(6)**

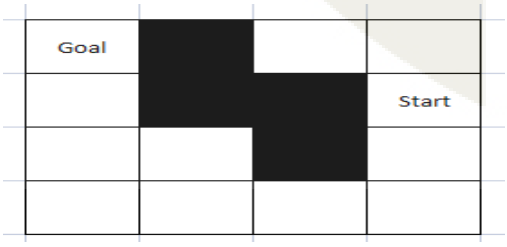
19. (a) Compare and contrast local and global Dynamic window approaches in obstacle avoidance. **(7)**

(b) Explain the concepts of floor finding Algorithm **(7)**

OR

20. (a) Illustrate the Incorporation of Neural network approach in Robot navigation? List its advantages **(6)**

(b) Make the robot to run from start position to goal position in the Following diagram using A* Algorithm **(8)**



TEACHING PLAN

No	Contents	No. of Lecture Hours (45 hrs)
Module-1 (Introduction to robotics) (8 hours)		
1.1	Introduction to robotics – Degrees of freedom - Robot types	1 hour
1.2	Manipulators- Anatomy of a robotic manipulator-links, joints, actuators, sensors, controller	1 hour
1.3	Robot configurations-PPP, RPP, RRP, RRR- Mobile robots- wheeled	1 hour
1.4	Legged robots, Aerial robots, underwater robots, surface water robots -	1 hour
1.5	Dynamic characteristics of robot- speed of motion, load carrying capacity & speed of response	1 hour
1.6	Introduction to End effectors - mechanical grippers, special tools, Magnetic grippers	1 hour
1.7	Vacuum grippers, adhesive grippers, Active and Passive grippers	1 hour
1.8	Ethics in robotics - 3 laws - applications of robots	1 hour
Module-2 (Sensors, Actuators and Control) (9 hours)		
2.1	Sensor classification- touch, force, proximity, vision sensors.	1 hour
2.2	Internal sensors-Position sensors, velocity sensors	1 hour
2.3	Acceleration sensors, Force sensors;	1 hour
2.4	External sensors-contact type, non-contact type	1 hour

2.5	Digital Camera - CCD camera - CMOS camera	1 hour
2.6	Omnidirectional cameras - Sensor characteristics	1 hour
2.7	Actuators - DC Motors - H-Bridge - Pulse Width Modulation	1 hour
2.8	Stepper Motors – Servos - Control - On-Off Control	1 hour
2.9	PID Control - Velocity Control and Position Control	1 hour
Module-3 (Robotic vision & Kinematics) (9 hours)		
3.1	Robot Vision: Sensing, Pre-processing, Segmentation, Description	1 hour
3.2	Recognition, Interpretation, Feature extraction -Camera sensor hardware interfacing	1 hour
3.3	Representation of Transformations - Representation of a Pure Translation - Pure Rotation about an Axis	1 hour
3.4	Combined Transformations - Transformations Relative to the Rotating Frame	1 hour
3.5	Basic understanding of Differential Drive Wheeled Mobile Robot - Car Like Wheeled Mobile Robot	1 hour
3.6	Kinematic model of a differential drive and a steered mobile robot.	1 hour
3.7	Degree of freedom and manoeuvrability, Degree of steerability	1 hour
3.8	Degree of mobility, Different wheel configurations	1 hour
3.9	Holonomic and Nonholonomic robots, Omnidirectional Wheeled Mobile Robots	1 hour
Module-4 (Localization and Mapping) (9 hours)		

4.1	Position and Orientation - Representing robot position, Basics of reactive navigation	1 hour
4.2	Robot Localization, Challenges in localization	1 hour
4.3	An error model for odometric position estimation	1 hour
4.4	Map Representation - Continuous representations - Decomposition strategies	1 hour
4.5	Current challenges in map representation, Probabilistic map-based localization (only Kalman method)	1 hour
4.6	Probabilistic map-based localization (only Kalman method)	1 hour
4.7	Autonomous map building, Simultaneous localization and mapping (SLAM) - Mathematical definition of SLAM	1 hour
4.8	Visual SLAM with a single camera - Graph-based SLAM	1 hour
4.9	Particle filter SLAM - Open challenges in SLAM	1 hour
Module-5 (Path Planning and Navigation) (10 hours)		
5.1	Path Planning- Graph search	1 hour
5.2	Deterministic graph search - breadth first search - depth first search- Dijkstra's algorithm	1 hour
5.3	A*, D* algorithms, Potential field based path planning	1.5 hour
5.4	Obstacle avoidance - Bug algorithm - Vector Field Histogram - Dynamic window approaches	1.5 hour

5.5	Navigation Architectures - Modularity for code reuse and sharing - Control localization - Techniques for decomposition	1 hour
5.6	Alternatives for navigation - Neural networks	1 hour
5.7	Processing the image - Training the neural network for navigation	1.5 hour
5.8	Training the neural network for navigation - Convolutional neural network robot control implementation	1.5 hour

****

CST 306	ALGORITHM ANALYSIS AND DESIGN	Category	L	T	P	Credit	Year of Introduction
		PCC	3	1	0	4	2019

Preamble:

The course introduces students to the design of computer algorithms, as well as analysis of algorithms. Algorithm design and analysis provide the theoretical backbone of computer science and are a must in the daily work of the successful programmer. The goal of this course is to provide a solid background in the design and analysis of the major classes of algorithms. At the end of the course students will be able to develop their own versions for a given computational task and to compare and contrast their performance.

Prerequisite:

Strong Foundation in Mathematics, Programming in C, Data Structures and Graph Theory.

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

CO#	CO
CO1	Analyze any given algorithm and express its time and space complexities in asymptotic notations. **(Cognitive Level: Apply)**
CO2	Derive recurrence equations and solve it using Iteration, Recurrence Tree, Substitution and Master's Method to compute time complexity of algorithms. **(Cognitive Level: Apply)**
CO3	Illustrate Graph traversal algorithms & applications and Advanced Data structures like AVL trees and Disjoint set operations. **(Cognitive Level: Apply)**
CO4	Demonstrate Divide-and-conquer, Greedy Strategy, Dynamic programming, Branch-and Bound and Backtracking algorithm design techniques (**Cognitive Level: Apply)**
CO5	Classify a problem as computationally tractable or intractable, and discuss strategies to address intractability **(Cognitive Level: Understand)**
CO6	Identify the suitable design strategy to solve a given problem. **(Cognitive Level: Analyze)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓								✓
CO2	✓	✓	✓	✓								✓
CO3	✓	✓	✓	✓								✓
CO4	✓	✓	✓	✓								✓
CO5	✓	✓										√
CO6	✓	✓	✓	✓								✓

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

Assessment Pattern

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

Analyze			
Evaluate			
Create			

Mark Distribution

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

Continuous Internal Evaluation Pattern:

Attendance	**10 marks**
Continuous Assessment Tests (Average of SeriesTests1&2)	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern:

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

End Semester Examination Pattern:

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

Syllabus

Module-1 (Introduction to Algorithm Analysis)

Characteristics of Algorithms, Criteria for Analysing Algorithms, Time and Space Complexity - Best, Worst and Average Case Complexities, Asymptotic Notations - Big-Oh (O), Big- Omega (Ω), Big-Theta (Θ), Little-oh (o) and Little- Omega (ω) and their properties. Classifying functions by their asymptotic growth rate, Time and Space Complexity Calculation of simple algorithms.

Analysis of Recursive Algorithms: Recurrence Equations, Solving Recurrence Equations – Iteration Method, Recursion Tree Method, Substitution method and Master's Theorem (Proof not required).

Module–2 (Advanced Data Structures and Graph Algorithms)

Self Balancing Tree - AVL Trees (Insertion and deletion operations with all rotations in detail, algorithms not expected); Disjoint Sets- Disjoint set operations, Union and find algorithms.

DFS and BFS traversals - Analysis, Strongly Connected Components of a Directed graph, Topological Sorting.

Module–3 (Divide & Conquer and Greedy Strategy)

The Control Abstraction of Divide and Conquer- 2-way Merge sort, Strassen's Algorithm for Matrix Multiplication-Analysis. The Control Abstraction of Greedy Strategy- Fractional Knapsack Problem, Minimum Cost Spanning Tree Computation- Kruskal's Algorithms - Analysis, Single Source Shortest Path Algorithm - Dijkstra's Algorithm-Analysis.

Module-4 (Dynamic Programming, Back Tracking and Branch & Bound))

The Control Abstraction- The Optimality Principle- Matrix Chain Multiplication-Analysis, All Pairs Shortest Path Algorithm - Floyd-Warshall Algorithm-Analysis. The Control Abstraction of Back Tracking – The N Queen's Problem. Branch and Bound Algorithm for Travelling Salesman Problem.

Module-5 (Introduction to Complexity Theory)

Tractable and Intractable Problems, Complexity Classes – P, NP, NP- Hard and NP-Complete Classes- NP Completeness proof of Clique Problem and Vertex Cover Problem- Approximation algorithms- Bin Packing, Graph Coloring. Randomized Algorithms (Definitions of Monte Carlo and Las Vegas algorithms), Randomized version of Quick Sort algorithm with analysis.

Text Books

1. T.H.Cormen, C.E.Leiserson, R.L.Rivest, C. Stein, Introduction to Algorithms, 2nd Edition, Prentice-Hall India (2001)
2. Ellis Horowitz, Sartaj Sahni, Sanguthevar Rajasekaran, "Fundamentals of Computer Algorithms", 2nd Edition, Orient Longman Universities Press (2008)

3. Sara Baase and Allen Van Gelder —Computer Algorithms, Introduction to Design and Analysis, 3rd Edition, Pearson Education (2009)

Reference Books

1. Jon Kleinberg, Eva Tardos, "Algorithm Design", First Edition, Pearson (2005)
2. Robert Sedgewick, Kevin Wayne, "Algorithms",4th Edition Pearson (2011)
3. GIlles Brassard, Paul Brately, "Fundamentals of Algorithmics", Pearson (1996)
4. Steven S. Skiena, "The Algorithm Design Manual", 2nd Edition, Springer(2008)

Course Level Assessment Questions

Course Outcome 1 (CO1):

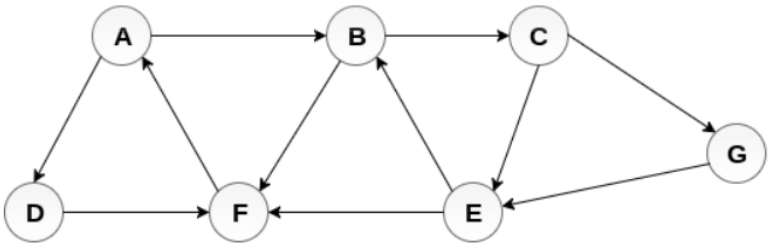
1. Is $2^{n+1} = O(2^n)$? Is $2^{2n} = O(2^n)$? Justify your answer.
2. What is the need of asymptotic analysis in calculating time complexity? What are the notations used for asymptotic analysis?
3. Calculate the time complexity for addition of two matrices.
4. Define time complexity and space complexity. Write an algorithm for adding n natural numbers and analyse the time and space requirements of the algorithm.

Course Outcome 2 (CO2):

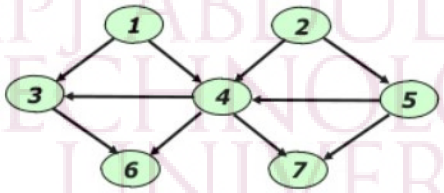
1. State Master's theorem for solving recurrences.
2. Solve the recurrence T(n) = 3T(n-2), using iteration method
3. State the conditions in recurrences where Master Theorem is not applicable.
4. Solve the following recurrence equations using Master's theorem.
   a) $T(n) = 8T(n/2) + 100\ n^2$
   b) $T(n) = 2T(n/2) + 10\ n$
5. Using Recursion Tree method, Solve T(n)= 2T(n/10)+ T(9n/10)+n. Assume constant time for small values of n.

Course Outcome 3 (CO3):

1. Explain the rotations performed for insertion in AVL tree with example.
2. Write down BFS algorithm and analyse the time complexity. Perform BFS traversal on the given graph starting from node A. If multiple node choices are available for next travel, choose the next node in alphabetical order.

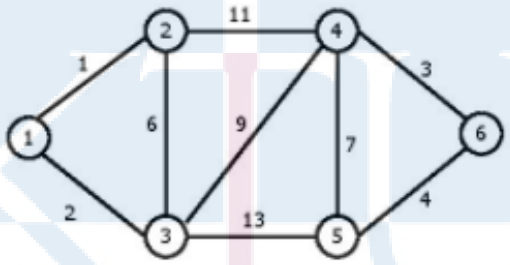


3. Find the minimum and maximum height of any AVL-tree with 7 nodes? Assume that the height of a tree with a single node is 0. (3)
4. Find any three topological orderings of the given graph.

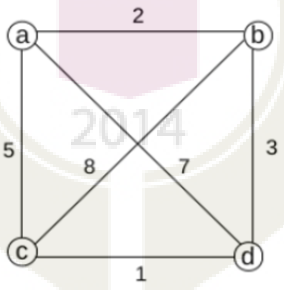


Course Outcome 4 (CO4):

1. Give the control abstraction for Divide and Conquer method.
2. Construct the minimum spanning tree for the given graph using Kruskal's algorithm. Analyse the complexity of the algorithm.



3. Compare Divide and Conquer and Dynamic programming methodologies
4. What is Principle of Optimality?
5. Define Travelling Salesman Problem (TSP). Apply branch and bound algorithm to solve TSP for the following graph, assuming the start city as 'a'. Draw the state space tree.



Course Outcome 5 (CO5):

1. Compare Tractable and Intractable Problems
2. With the help of suitable code sequence convince Vertex Cover Problem is an example of NP-Complete Problem

3. Explain Vertex Cover problem using an example. Suggest an algorithm for finding Vertex Cover of a graph.
4. Write short notes on approximation algorithms.
5. Compare Conventional quick sort algorithm and Randomized quicksort with the help of a suitable example?

Course Outcome 6 (CO6): (CO attainment through assignment only, not meant for examinations)

Choosing the best algorithm design strategy for a given problem after applying applicable design strategies – Sample Problems Given.

1. Finding the Smallest and Largest elements in an array of 'n' numbers
2. Fibonacci Sequence Generation.
3. Merge Sort
4. Travelling Sales Man Problem
5. 0/1 Knapsack Problem

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST 306

Course Name: Algorithm Analysis and Design

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

PART A

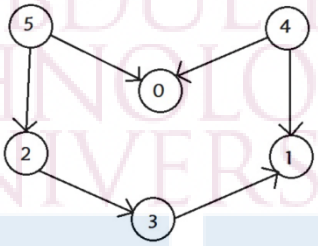
Answer All Questions. Each Question Carries 3 Marks

1. Define asymptotic notation? Arrange the following functions in increasing order of asymptotic growth rate.

n^3, 2^n, $\log n^3$, 2^{100}, $n^2 \log n$, n^n, $\log n$, $n^{0.3}$, $2^{\log n}$

2. State Master's Theorem. Find the solution to the following recurrence equations using Master's theorem.
   a) $T(n) = 8T(n/2) + 100\ n^2$
   b) $T(n) = 2T(n/2) + 10\ n$

3. Find any two topological ordering of the DAG given below.



4. Show the UNION operation using linked list representation of disjoint sets.

5. Write the control abstraction of greedy strategy to solve a problem.

6. Write an algorithm based on divide-and-conquer strategy to search an element in a given list. Assume that the elements of list are in sorted order.

7. List the sequence of steps to be followed in Dynamic Programming approach.

8. Illustrate how optimal substructure property could be maintained in Floyd-Warshall algorithm.

9. Differentiate between P and NP problems.

10. Specify the relevance of approximation algorithms.

(10x3=30)

Part B

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

11. (a) Define Big O, Big Ω and Big Θ Notation and illustrate them graphically. **(7)**

    (b) Solve the following recurrence equation using recursion tree method **(7)**

    $T(n) = T(n/3) + T(2n/3) + n$, where $n>1$

    $T(n) = 1$, Otherwise

OR

12. (a) Explain the iteration method for solving recurrences and solve the following recurrence equation using iteration method. **(7)**

$$T(n) = 3T(n/3) + n;\ T(1) = 1$$

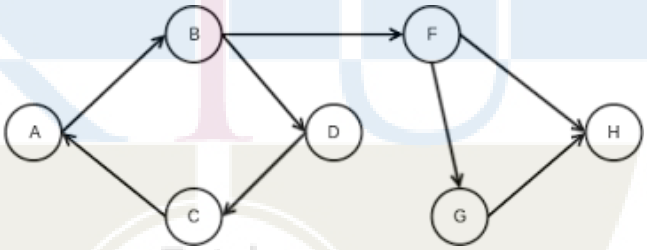
(b) Determine the time complexities of the following two functions fun1( ) and fun2( ). **(7)**

i)
```
int fun1(int n)
{
    if (n <= 1) return n;
    return 2*fun1(n-1);
}
```

ii)
```
int fun2 (int n)
{
    if (n <= 1) return n;
    return fun2 (n-1) + fun2 (n-1)
}
```

13. (a) Write DFS algorithm and analyse its time complexity. Illustrate the classification of edges in DFS traversal. **(7)**

(b) Find the strongly connected components of the digraph given below: **(7)**



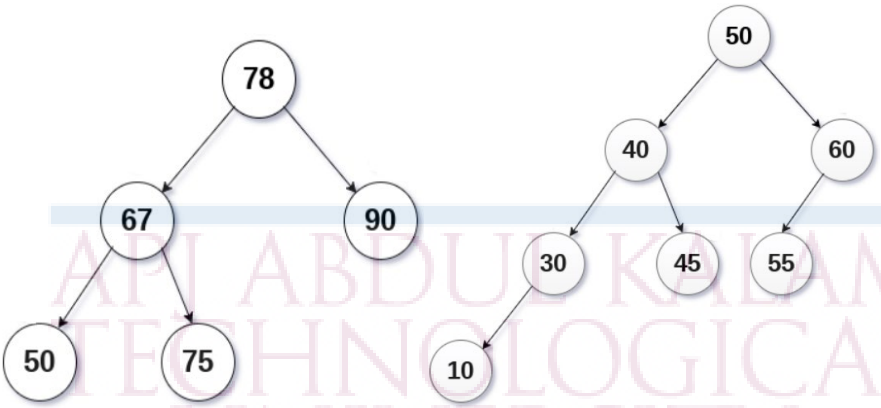
OR

14. (a) Illustrate the advantage of height balanced binary search trees over binary search trees? Explain various rotations in AVL trees with example. **(7)**

(b) Perform the following operations in the given AVL trees. **(7)**

i) Insert 70

ii) Delete 55



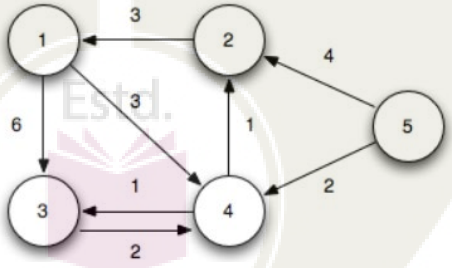
15. (a) State Fractional Knapsack Problem and write Greedy Algorithm for Fractional Knapsack Problem. **(7)**

(b) Find the optimal solution for the following Fractional Knapsack problem. Given the number of items(n) = 7, capacity of sack(m) = 15, W={2,3,5,7,1,4,1} and P = {10,5,15,7,6,18,3} **(7)**

OR

16. (a) Write and explain merge sort algorithm using divide and conquer strategy using the data {30, 19, 35, 3, 9, 46, 10}. Also analyse the time complexity. **(7)**

(b) Write the pseudo code for Dijkstra's algorithm. Compute the shortest distance from vertex 1 to all other vertices using Dijkstra's algorithm. **(7)**



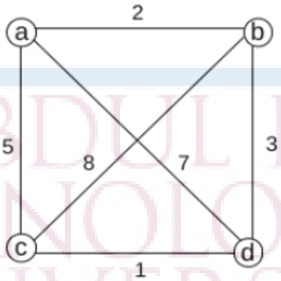
17. (a) Write Floyd-Warshall algorithm and analyse its complexity. **(5)**

(b) Write and explain the algorithm to find the optimal parenthesization of matrix chain product whose sequence of dimension is 4x10,10x3, 3x12,12x20. **(9)**

OR

18. (a) Explain the concept of Backtracking method using 4 Queens problem. **(7)**

(b) Define Travelling Salesman Problem (TSP). Apply branch and bound algorithm to solve TSP for the following graph, assuming the start city as 'a'. Draw the state space tree. **(7)**



19. (a) State bin packing problem? Explain the first fit decreasing strategy **(7)**

(b) Prove that the Clique problem is NP-Complete. **(7)**

OR

20. (a) Explain the need for randomized algorithms. Differentiate Las Vegas and Monte Carlo algorithms. **(6)**

(b) Explain randomized quicksort and analyse the expected running time of randomized quicksort with the help of a suitable example? **(9)**

Teaching Plan

No	Topic	No. of Hours (45 hrs)
Module -1 (Introduction to Algorithm Analysis) 9 hrs.		
1.1	Introduction to Algorithm Analysis: Characteristics of Algorithms.	1 hour
1.2	Criteria for Analysing Algorithms, Time and Space Complexity - Best, Worst and Average Case Complexities.	1 hour
1.3	Asymptotic Notations - Properties of Big-Oh (O), Big- Omega (Ω), Big-Theta (Θ), Little-Oh (o) and Little- Omega (ω).	1 hour
1.4	Illustration of Asymptotic Notations	1 hour

1.5	Classifying functions by their asymptotic growth rate	1 hour
1.6	Time and Space Complexity Calculation of algorithms/code segments.	1 hour
1.7	Analysis of Recursive Algorithms: Recurrence Equations, Solving Recurrence Equations – Iteration Method.	1 hour
1.8	Recursion Tree Method	1 hour
1.9	Substitution method and Master's Theorem and its Illustration.	1 hour
Module-2 (Advanced Data Structures and Graph Algorithms) 10 Hrs.		
2.1	Self Balancing Trees - Properties of AVL Trees, Rotations of AVL Trees	1 hour
2.2	AVL Trees Insertion and Illustration	1 hour
2.3	AVL Trees Deletion and Illustration	1 hour
2.4	Disjoint set operations.	1 hour
2.5	Union and find algorithms.	1 hour
2.6	Illustration of Union and find algorithms	1 hour
2.7	Graph Algorithms: BFS traversal, Analysis.	1 hour
2.8	DFS traversal, Analysis.	1 hour
2.9	Strongly connected components of a Directed graph.	1 hour
2.10	Topological Sorting.	1 hour
Module-3 (Divide & Conquer and Greedy Method) 8 Hrs		
3.1	Divide and Conquer: The Control Abstraction.	1 hour
3.2	2-way Merge Sort, Analysis.	1 hour
3.3	Strassen's Algorithm for Matrix Multiplication, Analysis	1 hour

3.4	Greedy Strategy: The Control Abstraction.	1 hour
3.5	Fractional Knapsack Problem.	1 hour
3.6	Minimum Cost Spanning Tree Computation- Kruskal's Algorithm, Analysis.	1 hour
3.7	Single Source Shortest Path Algorithm - Dijkstra's Algorithm	1 hour
3.8	Illustration of Dijkstra's Algorithm-Analysis.	1 hour
Module-4 (Dynamic Programming, Back Tracking and Branch and Bound) 8 Hrs.		
4.1	Dynamic Programming: The Control Abstraction, The Optimality Principle.	1 hour
4.2	Matrix Chain Multiplication-Analysis.	1 hour
4.3	Illustration of Matrix Chain Multiplication-Analysis.	1 hour
4.4	All Pairs Shortest Path Algorithm- Analysis and Illustration of Floyd-Warshall Algorithm.	1 hour
4.5	Back Tracking: The Control Abstraction .	1 hour
4.6	Back Tracking: The Control Abstraction – The N Queen's Problem.	1 hour
4.7	Branch and Bound:- Travelling salesman problem.	1 hour
4.8	Branch and Bound:- Travelling salesman problem.	1 hour
Module-5 (Introduction to Complexity Theory) 10 Hrs		
5.1	Introduction to Complexity Theory: Tractable and Intractable Problems.	1 hour
5.2	Complexity Classes – P, NP.	1 hour
5.3	NP- Hard and NP-Complete Problems.	1 hour
5.4	NP Completeness Proof of Clique Problem.	1 hour

5.5	NP Completeness Proof of Vertex Cover Problem.	1 hour
5.6	Approximation algorithms- Bin Packing Algorithm and Illustration.	1 hour
5.7	Graph Colouring Algorithm and Illustration.	1 hour
5.8	Randomized Algorithms (definitions of Monte Carlo and Las Vegas algorithms).	1 hour
5.9	Randomized Version of Quick Sort Algorithm with Analysis.	1 hour
5.10	Illustration of Randomized Version of Quick Sort Algorithm with Analysis.	1 hour

ADT342	DATA VISUALIZATION	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTIO
		PEC	2	1	0	3	2019

Preamble: The syllabus is prepared with the view of facilitating the learner to get an overview of data visualization. This course aims at providing fundamental knowledge in various data visualization techniques using R programming language and D3. It also deals with security aspects involved in data visualization. The learner will be able to understand the process and security aspects involved in data visualization and apply the tools in solving complex problems.

Prerequisite: Programming experience in any language and basic knowledge in R.

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

CO 1	Summarize the key techniques and theory used in visualization **(Cognitive Knowledge Level : Understand)**
CO 2	Design and use various methodologies present in data visualization. **(Cognitive Knowledge Level : Understand)**
CO 3	Employ appropriate processes and tools for data visualization. **(Cognitive Knowledge Level : Apply)**
CO 4	Use interactive data visualization to make inferences. **(Cognitive Knowledge Level : Apply)**
CO 5	Recognize the process involved and security issues present in data visualization. **(Cognitive Knowledge Level : Understand)**

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO1	✓											✓
CO2	✓	✓	✓									✓
CO3	✓	✓	✓	✓	✓							✓
CO4	✓	✓	✓	✓	✓							✓
CO5	✓	✓										✓

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

Assessment Pattern

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

Mark distribution

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

Continuous Internal Evaluation Pattern:

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

Internal Examination Pattern:
Each of the two internal examinations has to be conducted out of 50 marks

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

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

End Semester Examination Pattern:

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

SYLLABUS

Module 1 (Introduction to Data Visualization)
Introduction to Visualization – Need and purpose, External representation – Interactivity – Difficulty in Validation, Data Abstraction: Dataset types – Attribute types – Semantics, Task Abstraction – Analyze, Produce, Search, Query, Four levels of validation – Validation approaches – Validation examples. Marks and Channels. Data Visualization tools.

Module 2 (Arranging Spatial Data and Networks)
Arrange tables: Categorical regions – Spatial axis orientation – Spatial layout density, Arrange spatial data: Geometry – Scalar fields – Vector fields – Tensor fields. Arrange networks and trees: Connections, Matrix views – Containment, Map color: Color theory, Color maps and other channels.

Module 3 (Data Visualization using R)
Basic and Interactive Plots: scatter plot, interactive scatter plot, bar plot, line plot, interactive Gantt/timeline chart, Merging histograms, interactive bubble plot, waterfall plot, Heat Maps and Dendrograms: simple dendrogram, dendrograms with colors and labels, heat map, heat map with customized colors, three-dimensional heat map and a

stereo map, tree map. Maps: regional maps, choropleth maps, contour maps, maps with bubbles, Integrating text with maps, shapefiles, cartograms, Pie Chart and Its Alternatives, Adding the Third Dimension: 3D scatter plot, 3D pie chart, 3D histogram, 3D contour plot.

Module 4 (Interactive Data Visualization using D3)

Drawing with data: Drawing divs, SVG's, Making a bar chart, scatterplot – Scales - Axes – Updates, Transition and Motion – Modernizing the bar chart, Updating data, transitions, Interactivity – Layouts – Geomapping – Framework – D3.js, tableau.

Module 5 (Security Data Visualization)

Port scan visualization - Vulnerability assessment and exploitation - Firewall log visualization - Intrusion detection log visualization - Attacking and defending visualization systems – Creating security visualization system.

Text Books

1. Tamara Munzner, Visualization Analysis and Design, AK Peters Visualization Series, CRC Press, Nov. 2014
2. Atmajitsinh Gohil, "R Data Visualization Cookbook", PACKT, 2015.
3. Scott Murray, "Interactive data visualization for the web", O'Reilly Media, Inc., 2013.
4. Greg Conti, "Security Data Visualization: Graphical Techniques for Network Analysis", NoStarch Press Inc, 2007.

Reference Books

1. A Julie Steele and Noah Iliinsky, Designing Data Visualizations: Representing Informational Relationships, O'Relly.
2. Andy Kirk, Data Visualization: A Successful Design Process, PAKT.
3. Nathan Yau, "Data Points: Visualization that means something", Wiley, 2013.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1): Explain the four levels of validation in Data Visualization.

Course Outcome 2 (CO2): Discuss the different methods to arrange spatial data.

Course Outcome 3(CO3): Write sample code in R to generate a simple pie chart showing data on brain injury across different branches of the military:

Military Branch	Army	Navy	Air Force	Marines
No:	179718	41370	41914	44280

Also draw the resultant pie chart.

Course Outcome 4 (CO4): Given a dataset: [5, 10, 13, 19, 21, 25, 22, 18, 15, 13, 11, 12, 15, 20, 18, 17, 16, 18, 23, 25] to plot a bar graph. This dataset was later

modified as [11, 12, 15, 20, 18, 17, 16, 18, 23, 25, 5, 10, 13, 19, 21, 25, 22, 18, 15, 13]. Write the sample code in d3 to update the contents of a bar chart with new data values.

Course Outcome 5 (CO5): Explain Intrusion detection log visualization.

Model Question Paper

QPCODE: PAGES: 3

RegNo:_____

Name :_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: **ADT 342**

Course Name: Data Visualization

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

PART A

Answer all Questions. Each question carries 3 Marks

1. What is data visualization? Why do we use it? Illustrate the need of data visualization.
2. Why do data semantics and data types matter in data visualization?
3. Describe the HSL system.
4. Describe scatterplot with an example.
5. How is the waterfall plot constructed in R?
6. Compare chloropleth maps and cartograms.
7. Illustrate the use of ease() and delay() functions in transitions with an example.
8. Specify any three types of D3 layouts?
9. Describe port scan visualization.
10. What is meant by vulnerability assessment in visualization? (10x3=30)

Part B

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

11.a "Splitting the complex problem of visualization design into four cascading levels provides an analysis framework that lets you address different concerns separately". Explain the four cascading levels with a diagram? (10 marks)

11.b What are the threats to validity at each of the levels? (4 marks)

OR

12.a	Define marks and channels. Explain how visual channels control the appearance of marks. How are these visual channels and marks used for encoding various chart types?	(8 marks)
12.b	Illustrate Various data visualization tools.	(6 marks)
13.a	Differentiate between node-link diagrams and matrix views. Also specify the costs and benefits of each.	(8 marks)
13.b	Explain Treemaps and GrouseFlocks.	(6 marks)
	OR	
14.a	What is colour mapping? Explain the different types of colour maps.	(8 marks)
14.b	Explain scalar fields, vector fields and tensor fields.	(6 marks)
15.a	What is a dendrogram? Write the R code to construct a dendrogram.	(7 marks)
15.b	What is a pie chart? What are its limitations? Write the steps involved in its construction in R.	(7 marks)
	OR	
16.a	Why do we need a 3D scatter plot? Write the sample code to generate a 3D scatter plot in R.	(7 marks)
16.b	What are shape files? Why do we use them? Write the step-by-step procedure to construct a shape file.	(7 marks)
17.a	Given a data set = [5, 10, 13, 19, 21, 25, 22, 18, 15, 13,11, 12, 15, 20, 18, 17, 16, 18, 23, 25]; Write the D3 code to plot the given data set as bar chart with dual encoding of the data values in terms of both height and color. The data bars should have centered labels. Also plot the resultant bar graph for the given data set.	(7 marks)
17.b	Given another data set dataset = [[5, 20], [480, 90], [250, 50], [100, 33], [330,95], [410, 12], [475, 44], [25, 67], [85, 21], [220, 88]]; where [[]] indicate an array within another array. Plot this data set and specify the name of the plot obtained	(7 marks)
	OR	
18.a	What are named transitions? Explain with an example.	(7 marks)
18.b	What are tooltips? Explain with an example. What are the different types of tooltips?	(7 marks)
19.a	Write about firewall log visualization.	(7 marks)
19.b	Discuss in detail about intrusion detection log visualization.	(7 marks)
	OR	
20.a	Describe the concept of attacking and defending visualization systems.	(7 marks)
20.b	Describe about security visualization system.	(7 marks)

(14X5=70)

Teaching Plan

No	Contents	No. of Lecture Hours (35)
	Module - 1 (Introduction to Data Visualization)	**(6 hours)**
1.1	Introduction to Visualization – Need and purpose	1 hour
1.2	Data Abstraction: Dataset types	1 hour
1.3	Attribute types – Semantics	1 hour
1.4	Task Abstraction, Four levels of validation	1 hour
1.5	Validation approaches	1 hour
1.6	Data Visualization tools	1 hour
	Module - 2 (Arranging Spatial Data and Networks)	**(7 hours)**
2.1	Arrange tables: Categorical regions – Spatial axis orientation	1 hour
2.2	Spatial layout density	1 hour
2.3	Arrange spatial data: Geometry – Scalar fields	1 hour
2.4	Vector fields – Tensor fields	1 hour
2.5	Arrange networks and trees: Connections, Matrix views – Containment	1 hour
2.6	Map color: Color theory, Color maps and other channels	1 hour
2.7	Map color	1 hour
	Module - 3 (Data Visualization using R)	**(8 hours)**
3.1	Basic and Interactive Plots: scatter plot, interactive scatter plot	1 hour
3.2	Interactive Gantt/timeline chart, Merging histograms, interactive bubble plot, waterfall plot	1 hour
3.3	Heat Maps and Dendrograms : simple dendrogram, dendrograms with colors and labels, heat map	1 hour
3.4	heat map with customized colors, three-dimensional heat map and a stereo map, tree map	1 hour
3.5	Maps: regional maps, choropleth maps, contour maps	1 hour
3.6	maps with bubbles, Integrating text with maps, shape files, cartograms	1 hour
3.7	Pie Chart and Its Alternatives	1 hour
3.8	Adding the Third Dimension: 3D scatter plot, 3D pie chart, 3D histogram, 3D contour plot.	1 hour

Module- 4 (Interactive Data Visualization using D3)		**(8 hours)**
4.1	Drawing with data	1 hour
4.2	Scales	1 hour
4.3	Axes	1 hour
4.4	Updates, Transition and Motion – Modernizing the bar chart	1 hour
4.5	Updating data, transitions	1 hour
4.6	Interactivity	1 hour
4.7	Layouts	1 hour
4.8	Geomapping	1 hour
Module- 5 (Security Data Visualization)		**(6 hours)**
5.1	Port scan visualization	1 hour
5.2	Vulnerability assessment and exploitation	1 hour
5.3	Firewall log visualization	1 hour
5.4	Intrusion detection log visualization	1 hour
5.5	Attacking and defending visualization systems	1 hour
5.6	Creating security visualization system	1 hour

AIT312	RECOMMENDATION SYSTEM	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2020

Preamble: The course is prepared with the view of facilitating the learner to get an overview of recommender system. This course covers the concepts like Introduction to basic concepts and Recent developments, Collaborative Filtering, Content-based recommendation, Knowledge based recommendation, Hybrid approaches and Evaluating Recommender System. The course enables the learners to develop state-of-the-art recommender systems that automate a variety of choice-making strategies with the goal of providing affordable, personal, and high-quality recommendations

Prerequisite: Nil

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

CO 1	Describe the basic concepts of recommender systems **(Cognitive Knowledge Level : Understand)**
CO 2	Summarize the features of constraint based and case-based knowledge-based recommender systems **(Cognitive Knowledge Level : Understand)**
CO 3	Explain different hybridizing algorithms and illustrate them with suitable examples. **(Cognitive Knowledge Level : Understand)**
CO 4	Analyze the design issues in offline recommender evaluation **(Cognitive Knowledge Level : Apply)**
CO 5	Explain the features of attack-resistant recommender systems **(Cognitive Knowledge Level : Understand)**

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO6	PO 7	PO 8	PO 9	PO1 0	PO1 1	PO1 2
CO1	✓	✓										✓
CO2	✓	✓	✓									✓
CO3	✓	✓										✓
CO4	✓	✓	✓									✓
CO5	✓	✓										✓

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

Assessment Pattern

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

Mark Distribution

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

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

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

Continuous Assessment Assignment **15 marks**

Internal Examination Pattern:

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

End Semester Examination Pattern:

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

SYLLABUS

Module 1(Introduction to basic concepts and Recent developments)

Introduction to basic concepts and Recent developments , Collaborative recommendation -User-based nearest neighbor recommendations, Item-based nearest neighbour recommendation, Collaborative recommendation ratings , Model-based and preprocessing-based approaches , Recent practical approaches and systems Content-based recommendation - Content representation and content similarity Similarity-based retrieval and Other text classification methods

Module 2 (Knowledge-based recommendation)

Knowledge-based recommendation - Knowledge representation and reasoning , Constraints, Cases and similarities, Interacting with constraint-based recommenders - Defaults Dealing with unsatisfiable requirements and empty result set, Proposing repairs for unsatisfiable requirements, Ranking the items/utility-based recommendation, Interacting with case-based recommenders, Critiquing -Compound critiquing, Dynamic critiquing

Module 3 (Hybrid recommendation approaches)

Hybrid recommendation approaches - Opportunities for hybridization Recommendation paradigms, Hybridization designs, Monolithic hybridization design - Feature combination hybrids, Feature augmentation hybrids, Parallelized hybridization design - Mixed hybrids, Switching hybrids, Weighted hybrids, Pipelined hybridization design Cascade hybrids, Meta-level hybrids Limitations of hybridization strategies

Module 4 (Evaluating Recommender Systems)

Introduction - Evaluation Paradigms , User Studies , Online Evaluation Offline Evaluation with Historical Data Sets, General Goals of Evaluation Design - Accuracy, Coverage , Confidence and Trust , Novelty , General Goals of Evaluation Design - Serendipity ,Diversity , Robustness and Stability Scalability, Design Issues in Offline Recommender Evaluation - Case Study of the Netflix Prize Data Set , Segmenting the Ratings for Training and Testing - Hold-Out , Cross-Validation , Comparison with Classification, Accuracy Metrics in Offline Evaluation - Measuring the Accuracy of Ratings Prediction , RMSE versus MAE, Impact of the Long Tail, Evaluating Ranking via Correlation , Evaluating Ranking via Utility Evaluating Ranking via Receiver Operating Characteristic, Limitations of Evaluation Measures - Avoiding Evaluation Gaming

Module 5 (Attack-Resistant Recommender Systems)

Introduction Understanding the Trade-Offs in Attack Models - Quantifying Attack Impact Types of Attacks - Random Attack . Average Attack Bandwagon, Reverse Bandwagon Attack, Probe Attack Segment Attack, Effect of Base Recommendation Algorithm, Detecting Attacks on Recommender Systems - Individual Attack Profile Detection , Group Attack Profile Detection - Preprocessing Methods Online Methods Strategies for Robust Recommender Design - Preventing Automated Attacks with CAPTCHAs Using Social Trust. Designing Robust Recommendation Algorithms - Incorporating Clustering in Neighborhood Methods Fake Profile Detection during Recommendation Time Association - Based Algorithms

Text Books

1. Jannach D., Zanker M. and FelFering A., Recommender Systems: An Introduction, Cambridge University Press(2011)
2. C.C. Aggarwal, Recommender Systems: The Textbook, Springer, 2016.

Reference Books

1. F. Ricci, L Rokach, B. Shapira and P.B. Kantor, Recommender systems handbook, Springer 2010
2. Manouselis N., Drachsler H., Verbert K., Duval E., Recommender Systems For Learning, Springer (2013), 1st ed.

Course Level Assessment Questions

Course Outcome 1 (CO1): Discuss the cases in which content-based recommendations will not perform as well as collaborative filtering.

Course Outcome 2 (CO2): Analyze, in detail, different techniques available to support users in the interaction with constraint-based recommender applications.

Course Outcome 3(CO3): Explain about the feature combination and feature augmentation hybrid mechanisms.

Course Outcome 4 (CO4): With appropriate case study, explain the design issues in offline recommender evaluation .

Course Outcome 5 (CO5): Illustrate different methods used to detect attacks on existing recommender system.

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 3**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AIT312

Course Name: Recommendation System

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

PART A

Answer All Questions. Each Question Carries 3 Marks

1. List any two purposes of recommender systems.
2. Indicate the main idea of collaborative recommendation approaches.
3. Describe the two types of outputs generated with pure collaborative approaches that takes matrix of given user–item ratings as the only input .
4. Define case amplification. How it can be computed?
5. Explain the need of item-based nearest neighbor recommendation system.
6. Define classical constraint satisfaction problem (CSP).
7. Differentiate between parallelized hybridization design and pipelined hybridization design with suitable diagram.
8. What is meant by monolithic hybridization design?
9. Differentiate between internal validity and external validity.
10. Specify the effect of base recommendation algorithm. **(10x3=30)**

Part B

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

11. (a) Describe about user-based nearest neighbor recommendation system which deals with new items for which no ratings exist. **(7)**

(b) Explain about Rocchio's relevance feedback method **(7)**

OR

12. (a) Summarize the implicit and explicit rating mechanism in collaborative recommendation approaches. **(7)**

(b) Explain any two techniques that deal with data sparsity and the cold-start problem. **(7)**

13. (a) Explain ***QuickXPlain*** algorithm that calculates one conflict set at a time for a given set of constraints. **(7)**

(b) Which are the ways available to specify defaults? Explain how derived defaults can be determined. **(7)**

OR

14. (a) Explain about the ranking of items/utility-based recommendation. **(7)**

(b) Describe ***DynamicCritiquing*** algorithm. **(7)**

15. (a) Explain about feature combination hybrids. **(7)**

(b) Describe feature augmentation hybrid. **(7)**

OR

16. (a) Explain about different parallelized hybridization strategies. **(7)**

(b) Describe pipelined hybridization design methods. **(7)**

17. (a) Explain about offline and online evaluations in recommender systems. **(7)**

(b) Describe the general goals of evaluation design. **(7)**

OR

18. (a) Discuss about the design issues in offline recommender evaluation. Illustrate with a case study. **(7)**

(b) Explain about accuracy metrics in offline evaluation. **(7)**

19. (a) How do you quantify attack impact on recommender system? **(7)**

(b) Discuss about different attacks on recommender system. **(7)**

OR

20. (a) Discuss about different methods available to detect attacks on recommender system. **(7)**

(b) Explain how to design robust recommendation algorithms. **(7)**

Teaching Plan		
No	**Contents**	**No. of Lecture Hours (36 hrs)**
Module - 1 (Introduction to basic concepts)		**(7 hours)**
1.1	Introduction to basic concepts and Recent developments	1 hour
1.2	Collaborative recommendation :User-based nearest neighbor recommendation	1 hour
1.3	Collaborative recommendation : Item-based nearest neighbor recommendation	1 hour
1.4	Collaborative recommendation ratings	1 hour
1.5	Model-based and preprocessing-based approaches, Recent practical approaches and systems	1 hour
1.6	Content-based recommendation - Content representation and content similarity	1 hour
1.7	Similarity-based retrieval and Other text classification methods	1 hour
Module - 2 (Knowledge-based recommendation)		**(8 hours)**
2.1	Knowledge representation and reasoning , Constraints, Cases and similarities	1 hour
2.2	Interacting with constraint-based recommenders - Defaults	1 hour
2.3	Dealing with unsatisfiable requirements and empty result sets	1 hour
2.4	Proposing repairs for unsatisfiable requirements	1 hour
2.5	Ranking the items/utility-based recommendation	1 hour
2.6	Interacting with case-based recommenders Introduction	1 hour
2.7	Critiquing -Compound critiquing	1 hour
2.8	Dynamic critiquing	1 hour
Module- 3 (Hybrid recommendation)		**(7 hours)**
3.1	Hybrid recommendation approaches - Opportunities for hybridization Recommendation paradigms, Hybridization designs	1 hour
3.2	Monolithic hybridization design - Feature combination hybrids	1 hour
3.3	Feature augmentation hybrids	1 hour
3.4	Parallelized hybridization design -Mixed hybrids, Switching hybrids	1 hour
3.5	Weighted hybrids	1 hour

3.6	Pipelined hybridization design Cascade hybrids, Meta-level hybrids	1 hour
3.7	Limitations of hybridization strategies	1 hour
Module - 4 (Evaluating Recommender Systems)		**(8 hours)**
4.1	Introduction - Evaluation Paradigms , User Studies , Online Evaluation Offline Evaluation with Historical Data Sets	1 hour
4.2	Goals of Evaluation Design - Accuracy, Coverage , Confidence and Trust , Novelty	1 hour
4.3	General Goals of Evaluation Design - Serendipity ,Diversity , Robustness and Stability Scalability	1 hour
4.4	Design Issues in Offline Recommender Evaluation - Case Study of the Netflix Prize Data Set	1 hour
4.5	Design Issues in Offline Recommender Evaluation -Segmenting the Ratings for Training and Testing - Hold-Out , Cross-Validation , Comparison with Classification	1 hour
4.6	Accuracy Metrics in Offline Evaluation - Measuring the Accuracy of Ratings Prediction , RMSE versus MAE, Impact of the Long Tail	1 hour
4.7	Evaluating Ranking via Correlation , Evaluating Ranking via Utility Evaluating Ranking via Receiver Operating Characteristic	
4.8	Limitations of Evaluation Measures - Avoiding Evaluation Gaming	1 hour
Module- 5 (Attack-Resistant Recommender Systems)		**(6 hours)**
5.1	Introduction Understanding the Trade-Offs in Attack Models - Quantifying Attack Impact	1 hour
5.2	Types of Attacks - Random Attack . Average Attack Bandwagon	1 hour
5.3	Reverse Bandwagon Attack , Probe Attack Segment Attack , Effect of Base Recommendation Algorithm	1 hour
5.4	Detecting Attacks on Recommender Systems - Individual Attack Profile Detection ,Group Attack Profile Detection - Preprocessing Methods Online Methods	1 hour
5.5	Strategies for Robust Recommender Design - Preventing Automated Attacks with CAPTCHAs Using Social Trust . Designing Robust Recommendation Algorithms - Incorporating Clustering in Neighborhood Methods Fake Profile Detection during Recommendation Time	1 hour
5.6	Association-Based Algorithms	1 hour

AIT322	CONCEPTS IN COMPUTER GRAPHICS AND IMAGE PROCESSING	Category	L	T	P	Credit	Year of Introduction
		PEC	2	1	0	4	2019

Preamble: The purpose of this course is to make awareness about strong theoretical relationships between computer graphics and image processing. This course helps the learner to understand three-dimensional environment representation in a computer, transformation of 2D/3D objects, basic mathematical techniques and algorithms used to build useful applications, imaging, and image processing techniques. The study of computer graphics and image processing develops the ability to create image processing frameworks for different domains and develops algorithms for emerging display technologies.

Prerequisite: A sound knowledge of Mathematics and a programming language.

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

CO#	CO
CO1	Describe the working principles of graphics devices**(Cognitive Knowledge level: Understand)**
CO2	Illustrate line drawing, circle drawing and polygon filling algorithms(Cognitive Knowledge level: Apply)
CO3	Demonstrate geometric representations, transformations on 2D & 3D objects, clipping algorithms and projection algorithms(Cognitive Knowledge level: Apply)
CO4	Summarize visible surface detection methods(Cognitive Knowledge level: Understand)
CO5	Summarize the concepts of digital image representation, processing and demonstrate pixel relationships(Cognitive Knowledge level: Apply)
CO6	Solve image enhancement and segmentation problems using spatial domain techniques(Cognitive Knowledge level: Apply)

Mapping of course outcomes with program outcomes

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

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

Assessment Pattern

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

Analyze			
Evaluate			
Create			

Mark Distribution

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

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

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

Continuous Assessment Assignment **15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. 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 any5.

End Semester Examination Pattern:

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

SYLLABUS

Module – 1 (Basics of Computer graphics and Algorithms)

Basics of Computer Graphics and its applications. Video Display devices- Refresh Cathode Ray Tubes, Random Scan Displays and systems, Raster scan displays and systems. Line drawing algorithms- DDA, Bresenham's algorithm. Circle drawing algorithms- Midpoint Circle generation algorithm, Bresenham's algorithm.

Module – 2 (Filled Area Primitives and transformations)

Filled Area Primitives- Scan line polygon filling, Boundary filling and flood filling. Two dimensional transformations-Translation, Rotation, Scaling, Reflection and Shearing, Composite transformations, Matrix representations and homogeneous coordinates. Basic 3D transformations.

Module - 3 (Clipping and Projections)

Window to viewport transformation. Cohen Sutherland Line clipping algorithm.Sutherland Hodgeman Polygon clipping algorithm. Three-dimensional viewing pipeline. Projections-Parallel and Perspective projections. Visible surface detection algorithms- Depth buffer algorithm, Scan line algorithm.

Module - 4 (Fundamentals of Digital Image Processing)

Introduction to Image processing and applications. Image as 2D data. Image representation ingrayscale, Binary and Colour images. Fundamental steps in image processing. Components of image processing system.Coordinate conventions. Sampling and quantization. Spatial and Gray Level Resolution. Basic relationship between pixels– neighbourhood, adjacency, connectivity.

Module - 5 (Image Enhancement in Spatial Domain and Image Segmentation)

Basic gray level transformation functions- Log transformations, Power-Law transformations, Contrast stretching. Histogram equalization. Basics of spatial filtering - Smoothing spatial filter-Linear and nonlinear filters, and Sharpening spatial filters-Gradient and Laplacian.

Fundamentals of Image Segmentation.Thresholding-Basics of Intensity thresholding and Global Thresholding. Region based Approach- Region Growing, Region Splitting and Merging. Edge Detection - Edge Operators- Sobel and Prewitt.

Text Book

1. Donald Hearn and M. Pauline Baker, Computer Graphics, PHI, 2e, 1996
2. Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing. Pearson, 4e, 2017

References

1) William M. Newman and Robert F. Sproull, Principles of Interactive Computer Graphics. McGraw Hill, 2001
2) Zhigang Xiang and Roy Plastock, Computer Graphics (Schaum's outline Series), McGraw Hill, 2019.

3) David F. Rogers , Procedural Elements for Computer Graphics, Tata McGraw Hill,2001.
4) M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis, and Machine Vision, Thomson India Edition, 4e, 2017.

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Compare the working principle of raster scan systems and random scan systems.
2. How much time is spent scanning across each row of pixels during screen refresh on a raster system with resolution of 1280*1024 and a refresh rate of 60 frames per second?

Course Outcome 2 (CO2):

1. Rasterize the line with end points(2,3) and (5,8) using Bresenham's line drawing algorithm.
2. Explain how the 4-connected area filling approach differs from 8- connected area filling in boundary filling algorithm

Course Outcome 3 (CO3):

1. Rotate a triangle ABC 45 degree counter clockwise about the pivot point (10,3), where the position vector of the coordinate ABC is given as A(4,1), B(5,2) and C(4,3).
2. Given a clipping window A(20,20), B(60,20), C(60,40) and D(20,40). Using Cohen Sutherland algorithm, find the visible portion of the line segment joining the points P(40,80) and Q(120,30)

Course Outcome 4 (CO4):

1. Explain scan line algorithm for detecting visible surfaces in an object.

Course Outcome 5 (CO5):

1. Give an image representation model and describe how the representation changes in grayscale, binary and colour images.
2. Consider an image segment shown below.

   3 1 2 1 (q)

   2 2 0 2

   1 2 1 1

   (p) 1 0 1 2

   (a) Let V={0,1} and compute the length of the shortest 4-,8- and m- path between p and q. If a particular path does not exist between these two points , explain why?

   (b) Repeat for V={1,2}.

3. The spatial resolution of an image is given by 128 X 128.What is its storage requirements if it is represented by 64 gray levels?

Course Outcome 6 (CO6):

1. A skilled medical technician is charged with the job of inspecting a certain class of monochrome images generated by electronic microscope. To facilitate the inspection, the technician uses image processing aids. However when he examines the images he finds the following problems.
   (a) Presence of bright isolated dots that are not of interest.
   (b) Lack of sharpness
   (c) Poor contrast

   Identify the sequence of preprocessing steps that the technician may use to overcome the above mentioned problems and explain it.

2. A 4x4, 4 bits/pixel original image is given by

$$\begin{pmatrix} 10 & 12 & 8 & 9 \\ 10 & 12 & 12 & 14 \\ 12 & 13 & 10 & 9 \\ 14 & 12 & 10 & 12 \end{pmatrix}$$

   (a) Apply histogram equalisation to the image by rounding the resulting image pixels to integers
   (b) Sketch the histogram of the original image and the histogram-equalised image.

3. You have Sobel operator and Laplacian operator for edge detection. Which operator will you select for edge detection in the case of noisy image? Explain.**(Assignment)**

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AIT322

Course Name: Concepts in Computer Graphics and Image Processing

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

PART A

Answer All Questions. Each Question Carries 3 Marks

1. Justify the approach of using integer arithmetic in Bresenham's line drawing algorithm.

2. Consider a raster system with a resolution of 1024*1024. What is the size of the raster needed to store 4 bits per pixel? How much storage is needed if 8 bits per pixel are to be stored?

3. Show that two successive reflections about either of the coordinate axes is equivalent to a single rotation about the coordinate origin.

4. Determine a sequence of basic transformations that are equivalent to the x-direction shearing matrix.

5. Find the window to viewport normalization transformation with window lower left corner at (1,1) and upper right corner at (2,6).

6. Find the orthographic projection of a unit cube onto the x=0, y=0 and z=0 plane.

7. Define Sampling and Quantization of an image.

8. Give any three applications of digital image processing.

9. A captured image appears very dark because of wrong lens aperture setting. Describe an enhancement technique which is appropriate to enhance such an image.

10. Suggest an approach of thresholding that should be used in case of uniform illumination. **(10x3=30)**

Part B

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

11. (a) Write Midpoint circle drawing algorithm and use it to plot a circle with radius=20 and center is (50,30). **(10)**

(b) Draw the architecture of raster scan display systems and explain its working principle. **(4)**

OR

12. (a) Derive the initial decision parameter of Bresenham's line drawing algorithm and use the algorithm to rasterize a line with endpoints (2,2) and (10,10). **(10)**

(b) Explain the working principle of color CRT monitors with suitable illustrations. **(4)**

13. (a) Compare boundary fill algorithm and flood fill algorithm. **(5)**

(b) Reflect a triangle ABC about the line 3x-4y+8=0. The position vector of the coordinate ABC is given as A(4,1), B(5,2) and C(4,3). **(9)**

OR

14. (a) Explain the need of using vanishing points in projections. **(4)**

(b) Explain Cohen-Sutherland line clipping algorithm. Use the algorithm to clip line P1(70, 20) and P2(100,10) against a window lower left hand corner (50,10) and upper right hand corner (80,40). **(10)**

15. (a) Describe Sutherland Hodegman polygon clipping algorithm and what are its limitations. **(7)**

(b) Explain how visible surfaces can be detected using depth buffer algorithm. **(7)**

OR

16. (a) Describe Sutherland Hodegman polygon clipping algorithm and what are its limitations. **(7)**

(b) Explain how visible surfaces can be detected using depth buffer algorithm. **(7)**

17. (a) Explain the components of an image processing system with suitable diagram **(9)**

(b) Define Resolution of an image. Explain the spatial and gray level resolution of an image with an example. **(5)**

OR

18. (a) Define 4-adjacency, 8 adjacency and m-adjacency. Consider the image segment shown. **(7)**

```
          4  2  3  2  (q)
   3  3  1  3
   2  3  2  2
         (p)  2  1  2  3
```

Let V={1,2} and compute the length of the shortest 4- ,8- and m- path between p and q. If a particular path does not exist between these two points, explain why?

(b) Using any one application, explain the steps involved in image processing. **(7)**

19. (a) A 5x5 image patch is shown below. Compute the value of the marked pixel if it is smoothened by a 3x3 average filterand median filter. **(4)**

$$f(m,n)=\begin{pmatrix} 0 & 1 & 2 & 3 & 2 \\ 5 & 6 & 7 & 8 & 4 \\ 4 & 3 & \textcircled{2} & 1 & 2 \\ 8 & 7 & 6 & 5 & 3 \\ 1 & 5 & 3 & 7 & 6 \end{pmatrix}$$

(b) Define Image segmentation and describe in detail method of edge and region **(10)**

based segmentation technique.

OR

20. (a) Distinguish between smoothing and sharpening filters in terms of **(10)**

(i) Functionality
(ii) Types
(iii) Applications
(iv) Mask Coefficients

(b) Describe how an image is segmented using split and merge technique in association with the region adjacency graph. **(8)**

TEACHING PLAN

No	Contents	No of Lecture Hrs (36 hrs)
Module – 1 (Basics of Computer Graphics and Algorithms) (8 hrs)		
1.1	Basics of Computer Graphics and applications	1 hour
1.2	Refresh Cathode Ray Tubes	1 hour
1.3	Random Scan Displays and systems,Raster scan displays and systems	1 hour
1.4	DDA Line drawing Algorithm	1 hour
1.5	Bresenham's line drawing algorithm	1 hour
1.6	Midpoint Circle generation algorithm	1 hour
1.7	Bresenham's Circle generation algorithm	1 hour
1.8	Illustration of line drawing and circle drawing algorithms	1 hour
Module - 2 (Filled Area Primitives and transformations) (8 hrs)		
2.1	Scan line polygon filling	1 hour
2.2	Boundary filling and flood filling	1 hour
2.3	Basic 2D transformations-Translation, Rotation and Scaling	1 hour

2.4	Reflection and Shearing	1 hour
2.5	Composite transformations	1 hour
2.6	Matrix representations and homogeneous coordinates	1 hour
2.7	Basic 3D transformation-Translation and scaling	1 hour
2.8	Basic 3D transformation-Rotation	1 hour
Module - 3 (Clipping and Projections) (7 hrs)		
3.1	Window to viewport transformation	1 hour
3.2	Cohen Sutherland Line clipping algorithm	1 hour
3.3	Sutherland Hodgeman Polygon clipping algorithm	1 hour
3.4	Practice problems on Clipping algorithms	1 hour
3.5	Three-dimensional viewing pipeline, Projections-Parallel projections,Perspective projections	1 hour
3.6	Visible surface detection algorithms- Depth buffer algorithm	1 hour
3.7	Scan line visible surface detection algorithm	1 hour
Module - 4 (Fundamentals of Digital Image Processing) (6 hrs)		
4.1	Introduction to Image processing-Image as a 2D data, Image representation-Gray scale, Binary and Colour images.	1 hour
4.2	Fundamental steps in image processing and applications	1 hour
4.3	Components of image processing system	1 hour
4.4	Coordinate conventions, Sampling and quantization, Spatial and Gray Level Resolution	1 hour
4.5	Basic relationship between pixels – neighbourhood, adjacency, connectivity	1 hour
4.6	Illustration of basic relationship between pixels– neighbourhood, adjacency, connectivity	1 hour

Module - 5 (Image Enhancement in spatial domain and Image Segmentation) (7 hrs)		
5.1	Basic gray level transformation functions- Log transformations, Power law transformation, Contrast stretching	1 hour
5.2	Histogram equalization with illustration	1 hour
5.3	Basics of spatial filtering, Smoothing spatial filter- Linear and nonlinear filters	1 hour
5.4	Sharpening spatial filtering-Gradient filter mask, Laplacian Filter Mask	1 hour
5.5	Fundamentals of Image Segmentation, Basics of Intensity thresholding, Basic Global Thresholding	1 hour
5.6	Region Based Approach- Region Growing, Region Splitting and Merging	1 hour
5.7	Basics of Edge Detection- Sobel and Prewitt edge detection masks	1 hour

AIT352	ARTIFICIAL NEURAL NETWORKS TECHNIQUES	Category	L	T	P	Credit	Year of Introduction
		PEC	2	1	0	3	2019

Preamble: This course enables the learners to understand the fundamental concepts regarding Artificial Neural networks. The course covers basic analogy between ANN and human brain, the basic learning laws, fundamental ANN algorithms, Back Propagation Feed Forward Network, Self Organising Maps, RBF net, BAM and ART networks. This course enables the students to apply techniques and methods to solve real-world problems involving the application of ANN.

Prerequisite: Nil.

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

CO1	Summarize the basic concepts and the learning rules of ANN. **(Cognitive Knowledge Level: Understand)**
CO2	Utilize the fundamental learning algorithms namely, Mc-Culloch Pitts, Hebb Perceptron and Adaline to solve real world problems.**(Cognitive Knowledge Level: Apply)**
CO3	Implement Back propagation learning algorithm, Generic Radial Basis Function network. **(Cognitive Knowledge Level: Apply)**
CO4	Demonstrate Self Organizing Maps and Adaptive Resonance Theory.(**Cognitive Knowledge Level: Understand)**
CO5	Implement training algorithms for pattern association. **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

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

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

Assessment Pattern

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

Mark Distribution

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

Continuous Internal Evaluation Pattern:

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

Internal Examination Pattern

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

End Semester Examination Pattern:

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

SYLLABUS

Module – 1 (Basics of Artificial Neural Network and Learning Methods)

Characteristics of the human brain, Neurons, Introduction to Artificial Neural Networks, Terminology, Models of ANN, Topology, Network Architectures, Knowledge Representation, Learning Process, Learning Tasks. Categories of learning - Hebbian learning, Perceptron Learning Rule, Delta Learning Rule, Generalized Delta Learning Rule, Competitive learning, Error-correction learning, Reinforcement learning, Stability and Convergence.

Module – 2(Basic ANN Models)

McCulloch-Pitts Neuron, Architecture, Algorithm and Applications. Biases and Thresholds, Linear Separability. Hebb Net - Algorithm, Applications. Perceptron - Architecture, Algorithm, Applications. Perceptron Learning Rule Convergence Theorem. Adaline - Architecture, Algorithm, Applications.

Module - 3 (Multilayer Perceptrons)

Multi-Layered network architecture, Back propagation Algorithm, Applications, XOR problem, Replacing and Modifying Back propagation Algorithms Using Heuristics.

Cover's Theorem on the Separability of patterns, The Interpolation Problem, Radial Basis Function Networks, Comparison of MLP and RBF Networks(Theory only).

Module – 4 (SOMs and ART Networks)

Self-organizing maps - Building, Training, Evaluating, Interpreting and Visualizing a Self-organizing Map. Applications of Self Organizing Maps.

Adaptive Resonance Theory -Stability Plasticity Dilemma, ART-1-Architeture, Algorithm, Applications. ART-2 – Architeture, Algorithm, Applications.

Module – 5 (Training Algorithms for Pattern Association)

Introduction, Hetero associative neural network- Architecture, Applications. Auto Associative Net -Architecture, Applications. Iterative Auto Associative Net – Architecture, Applications. Discrete Hopfield Network. Bidirectional Auto-associative Memory – Architecture, Applications.

Text Books

1. Simon Haykin, "Neural Networks, A comprehensive Foundation"(2nd edition), Pearson Education (Module - 4)
2. Laurene Faucett, "Fundamentals of Artificial Neural Networks, architecture algorithm and applications"(Modules – 2,3,5)
3. Yegnanarayana, "Artificial Neural Networks", Phi Learning (Module -1)

Reference Books

1. Christopher M Bishop," Neural networks for Pattern Recognition
2. Mohammad H Hassoun, "Fundamentals of Artificial Neural Networks"

Course Level Assessment Questions

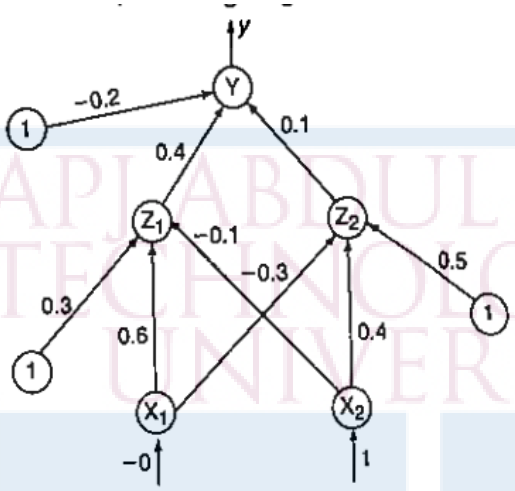
Course Outcome1 (CO1):

1. What are the different types of competitive learning?
2. Demonstrate the significance of different Activation functions.
3. Explain the terms cell body, axon, synapse, dendrite and neuron with reference to abiological neural network.
4. Illustrate examples of pattern recognition tasks to demonstrate the superiority of the biological neural network over a conventional computer system. (Assignment Question)

Course Outcome 2 (CO2):

1. How is training adopted in Adaline network and state the characteristics of weighted interconnections in Adaline .
2. How is the linear separability concept implemented using Perceptron Network training?
3. Implement NAND logical function using Perceptron Network in Python language(Assignment Question)

Course Outcome 3(CO3):



1. Find the new weights of Back propagation net shown in the figure for the input pattern (0,-1) and the target output 1, Use 0.25 as learning rate.
2. Why is gradient descent method adopted to minimize error? Explain in relation to Back propagation of error phase of BPNN?
3. Implement RBF network using Python language. (Assignment Question)

Course Outcome 4(CO4): .

1. Design an ART1 used to cluster four vectors with low vigilance. The values and description of the parameters are given in the table. Cluster the vectors, (1,1,0,0), (0,0,0,1), (1,0,0,0), (0,0,1,1) in at most three clusters.

n=4	Number of components in the input vector
m=3	It was an excellent game.
P=0.4	Vigilence parameter
L=2	Parameter used in update of bottom-up weights
$b_{ij}(0)=1/n+1$	Initial bottom-up weights
$t_{ij}(0)=1$	Initial top-down weights

2. Use NeuPy library of Python to implement Adaptive Resonance Theory (ART1) Network for binary data clustering.
3. Implement Self Organizing Map in Python to demonstrate how does the grid automatically arrange, using colour patterns and evaluate the effect of Learning Rate and Radius. (Assignment Question)

Course Outcome 5(CO5):

1. Compare and contrast auto associative and hetero associative networks with examples.

2. Implement Bidirectional Associative Memory using Python without using specific libraries. (Assignment Question)

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AIT352

Course Name: Artificial Neural Networks Techniques

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

PART A

Answer All Questions. Each Question Carries 3 Marks

1.	What are the main differences among the three models of artificial neurons, namely, McCulloch-Pitts, Perceptron and Adaline?	**3**
2.	Compare the stability and convergence of ANN	**3**
3.	Design a Mc-Culloch Pitts neural network to implement AND function.	**3**
4.	Define Perceptron Learning Rule Convergence theorem.	**3**
5.	What is the significance of momentum factor in backpropagation learning?	**3**
6.	Compare RBF network and Multilayer Perceptron network.	**3**
7.	Illustrate the feature mapping models.	**3**
8.	What is the significance of 'resonance' in ART network?	**3**

9.	Explain the hebb rule for pattern association		**3**
10.	Interpret cross talk and perfect recall using suitable examples		**3**

Part B

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

11.	(a)	Describe any four attractive features of the biological neural network that make it superior to the most sophisticated Artificial Intelligence computer system for pattern recognition tasks.	**(8)**
	(b)	Compare LMS, Perceptron and Delta learning laws.	**(6)**

OR

12.	(a)	Compare the performance of a computer and that of a biological neural network in terms of speed of processing, size and complexity, storage, fault tolerance and control mechanism.	**(8)**
	(b)	What is reinforcement learning? In what way it is different from supervised learning?	**(6)**
13.	(a)	Explain Hebb net algorithm and implement logical AND function using bipolar inputs.	**(4)**
	(b)	Use Adaline nerwork to train AND NOT function with bipolar inputs and targets. Perform one epoch of training.	**(10)**

OR

14.	(a)	Using the Perceptron Learning rule find the weights required to perform the following classifications. Vectors (1,1,1,1) and (-1,1,-1,-1) are members of the class and hence target value 1; vectors (1,1,1,-1) and (1, -1, -1, 1) are not the members of the class and hence target value -1. Use learning rate of 1 and starting value of weights as 0, test the response of the net.	**(10)**
	(b)	XOR function is non·linearly separable by a single decision boundary line. Justify.	**(4)**
15.	(a)	Analyse Cover's theorem based on XOR problem.	**(10)**
	(b)	Explain the learning factors of of Back propagation network algorithm	**(4)**

OR

16. (a) Relate Hidden layer and Output layer error terms with back propagation of error term phase in Back Propagation Network algorithm. **(10)**

(b) Explain the architecture and algorithm of RBF network . **(8)**

17. (a) Explain the statistical properties exhibited by SOM after convergence. **(10)**

(b) Interpret stability-plasticity dilemma in relation with ART network. **(4)**

OR

18. (a) Show the architecture of Kohonen's Self Organising Map and demonstrate the competitive process in Kohonen's self organising Map. **(8)**

(b) Explain the basic architecture of ART-2 and its algorithm. **(6)**

19. (a) Describe the architecture and algorithm of Discrete Bidirectional Associative Memory **(5)**

(b) Use the Hebb rule to store the vetors (1,1,1,1)and (1,1,-1,-1) in an auto associative neural net. **i.(9)**

a. Fnd the weight matrix(Do not set the diagonal terms to zero)

b. Test the net, using the following vectors as input

i. (1,1,1,1)

ii. (1,1,-1,-1)

iii. (1,1,1,0)

Repeat parts a and b with diagonal weight matrix set to zero. Identify the differences in the response.

OR

20. (a) Design a BAM net to associate the letters"A" and "C" given in bipolar 5X3 vectors to the bipolar codes (-1,1) and (1,1) respectively. **(10)**

(b) Compare Iterative Autoassocative with Discrete Hopfield Net. **(4)**

TEACHING PLAN

No	Contents	No of Lecture Hrs: 35
Module -1 : Basics of Artificial Neural Network and Learning methods (7 hours)		
1.1	Introduction to Neural Network, The human brain - Characteristics of Neural Network.	1
1.2	Artificial Neural Network - Terminology, Models of a neuron, Topology	1
1.3	Network architectures, Knowledge representation.	1
1.4	Learning Process, Learning tasks.Categories of learning- Hebbian learning, Competitive learning.	1
1.5	Error-correction learning.	1
1.6	Reinforcement learning.	1
1.7	Stability and Convergence.	1
Module - 2 : Basic ANN Models(7 hours)		
2.1	McCulloch-Pitts Neuron - Architecture, Algorithm and Applications.	1
2.2	Biases and thresholds, Linear separability.	1
2.3	Hebb net - Algorithm , Applications	1
2.4	Perceptron -Architecture, Algorithm	1
2.5	Perceptron -Applications, Perceptron learning rule convergence theorem.	1
2.6	Perceptron learning rule convergence theorem. Adaline - Architecture, Algorithm	1
2.7	Adaline - Applications	1
Module 3 : Multilayer Perceptrons (7 hours)		
3.1	Multilayered Feed Forward Network Architecture,	1

3.2	Back propagation algorithm, Activation functions, Rate of learning, Stopping criteria	1
3.3	Applications, XOR problem, Heuristics for making the Back propagation algorthm perform better.	1
3.4	Cover's Theorem on the separability of patterns.	1
3.5	Cover's Theorem on the separability of patterns, XOR problem.	1
3.6	The interpolation problem, Radial Basis Function networks.	1
3.7	The interpolation problem, Radial Basis function networks, Comparison of RBF network and Multi-Layer perceptrons.	1
Module 4 : SOMs and ART networks (7 hours)		
4.1	Two basic feature mapping methods.	1
4.2	Self Organizing Map, Competitive process, Cooperative process, Adaptive process.	1
4.3	Properties of the feature map.	1
4.4	Stability Plasticity Dilemma, ART-1-Architeture**.**	1
4.5	ART-1 - Algorithm, Applications.	1
4.6	ART-2 - Architeture-Algorithm	1
4.7	ART-2 - Applications.	1
Module 5 : Training Algorithms for pattern Association (7 hours)		
5.1	Introduction, Hebb rule for pattern association, Delta rule for pattern association	1
5.2	Hetero Associative Neural Network-Architecture , Applications,	1
5.3	Auto-associative Net - Architecture, Algorithm, Applications, Storage capacity.	1
5.4	Iterative Auto Associative Net - Architecture, Applications	1
5.5	Discrete Hopfield network - Architecture, Algorithm, Applications**.**	1
5.6	Bidirectional Auto-associative Memory-Architecture, Algorithm.	1
5.7	Bidirectional Auto-associative Memory – Applications.	1

AIT362	PROGRAMMING IN R	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PEC	2	1	0	3	2019

Preamble: The objective of this course is to enable the learner to make use of R Programming language to perform analysis and extraction of information from data irrespective of the quantity. It encompasses the R programming environment, syntax, data representations, data processing, statistical analysis and visualization. This course facilitates the learner to develop modular software solutions to perform statistical analysis and data extraction.

Prerequisite: Fundamental concepts in programming in C and Probability and Statistical Modeling

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

CO 1	Illustrate uses of conditional and iterative statements in R programs. (Cognitive Knowledge level: Apply)
CO 2	Write, test and debug R programs (Cognitive Knowledge level: Apply)
CO 3	Illustrate the use of Probability distributions and basic statistical functions. (Cognitive Knowledge level: Apply)
CO 4	Visualize different types of data (Cognitive Knowledge level: Apply)
CO 5	Comprehend regression modeling using R (Cognitive Knowledge level: Understand)

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO6	PO 7	PO 8	PO 9	PO10	PO11	PO12
CO1	✓	✓	✓		✓							✓
CO2	✓	✓	✓		✓							✓
CO3	✓	✓	✓	✓	✓							✓
CO4	✓	✓	✓	✓	✓							✓
CO5	✓	✓			✓							✓

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

Assessment Pattern

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

Mark distribution

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

Continuous Internal Evaluation Pattern:

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

Internal Examination Pattern:

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

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

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

End Semester Examination Pattern:

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

SYLLABUS

Module -1 (Introduction to R)

The R Environment - Command Line Interface and Batch processing, R Packages, Variables, Data Types, Vectors- vector operations and factor vectors, List- operations, Data Frames, Matrices and arrays, Control Statements- Branching and looping - For loops, While loops, Controlling loops. Functions- Function as arguments, Named arguments

Module -2(Reading and writing data)

Importing data from Text files and other software, Exporting data, importing data from databases- Database Connection packages, Missing Data - NA, NULL

Combining data sets, Transformations, Binning Data, Subsets, summarizing functions. Data Cleaning, Finding and removing Duplicates, Sorting.

Module -3 (Statistics with R)

Analyzing Data, Summary statistics, Statistical Tests- Continuous Data, Discrete Data, Power tests, Common distributions- type arguments. Probability distributions, Normal distributions

Module -4(Data Visualization)

R Graphics- Overview, Customizing Charts, Graphical parameters, Basic Graphics functions, Lattice Graphics - Lattice functions, Customizing Lattice Graphics, Ggplot.

Module - 5 (Regression Models)

Building linear models - model fitting, Predict values using models, Analyzing the fit, Refining the model, Regression- types, Unusual observation and corrective measures,

Comparison of models, Generalized linear models - Logistic Regression, Poisson Regression, Nonlinear least squares

Text Book

1. Joseph Adler, " R in a Nutshell", Second edition,O'reilly,2012

Reference Books

1. Jared P Lander, R for Everyone- Advanced analytics and graphics, Addison Wesley data analytics series, Pearson
2. Norman matloff, The art of R programming, A Tour of Statistical, Software Design, O'reilly
3. Robert Kabacoff, R in action, Data analysis and graphics with R, Manning
4. Garret Grolemund, Hands-on programming with R, Write your own functions and simulations, O'reilly

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. What is Coercion? How is it done in R?
2. Write a program to find the factorial of a number.
3. Write a program to compute roots of a quadratic equation.

Course Outcome 2 (CO2):

1. Write a program to read data from a table 'table123' in a database named 'db123' and display the values .
2. Explain Data cleaning in R
3. How missing data is handled in R?

Course Outcome 3(CO3):

1. Explain summary function in R
2. Illustrate how statistical testing is performed in R
3. Describe about probability distributions.

Course Outcome 4 (CO4):

1. Illustrate the use of ggplot() and various data visualization tools using appropriate datasets

Course Outcome 5 (CO5):

1. Illustrate the steps to predict the weight of a person when his height is unknown using linear regression for the data given below.

Height	151	174	138	186	128	136	179	163	152	130
Weight	63	81	56	91	47	57	76	72	62	48

Model Question Paper

QP CODE: PAGES:3

Reg No:_____

Name :_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: **AIT 362**

Course Name: Programming in R

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

PART A

Answer all Questions. Each question carries 3 Marks

1. Write a R program to add element "23" to the vector (24,56,67) in the second position.
2. Discuss the general list operations in R with example.
3. Calculate the cumulative sum and cumulative product for the given data 23, 1, 7,2,8,10, 17 using R Program.
4. Explain aggregate function in R.
5. List the applications of R programming.
6. Illustrate summary function.
7. List any three graphics functions.
8. Explain Lattice function.
9. Suppose that you have a dataset D1 and you design a linear regression model of degree 3 polynomial and you found that the training and testing error is "0" or in other terms it perfectly fits the data. What will happen when you fit a degree 2 polynomial in linear regression?
10. Explain logistic regression function in R.

(10x3=30)

Part B

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

11.a Write a R program to extract every nth element from a vector. (7 marks)

11.b Find the Nth highest value of a vector in R. (7 marks)

OR

12.a Write a R program to create a data frame using two given vectors and display the duplicate elements and unique rows of the said data frame. (7 marks)

12.b Write a R program to compare two data frames to find the row(s) in the first data frame that are not present in the second data frame. (7 marks)

13.a Write a R program to call the (built-in) dataset air quality. Remove the variables 'Solar.R' and 'Wind' and display the data frame. (7 marks)

13.b Illustrate transformation functions in R. (7 marks)

OR

14.a Write a R program to write the following data to a CSV file. (7 marks)

	Country	Population_1_july_2018	Population_1_july_2019	change_in_percents
1	China	1,427,647,786	1,433,783,686	+0.43%
2	India	1,352,642,280	1,366,417,754	+1.02%
3	United States	327,096,265	329,064,917	+0.60%
4	Indonesia	267,670,543	270,625,568	+1.10%
5	Pakistan	212,228,286	216,565,318	+2.04%

14.b Given a file "auto.csv" of automobile data with the fields index, company, body-style, wheel-base, length, engine-type, num-of-cylinders, horsepower, average-mileage, and price, write R program to print total cars of all companies, Find the average mileage of all companies. (7 marks)

15.a Write a note on data analysis using R. (7 marks)

15.b Explain how statistical test are performed using R functions. (7 marks)

OR

16.a Write R code to generate the probability distribution table for number of successes from a binomial distribution where n=5 and probability of success in each trial is 0.25. (7 marks)

16.b Fit a Poisson distribution with the following data using the following data (7 marks)

X	0	1	2	3	4	5
F	142	156	69	27	5	1

OR

17 Given the sales information of a company as CSV file with the following, fields month_number, face cream, facewash, toothpaste, bathingsoap, shampoo, moisturizer, total_units, total_profit. Write R codes to visualize the data as follows:

a) Toothpaste sales data of each month and show it using a scatter plot. (7 marks)

b) Calculate total sale data for last year for each product and show it using a Pie chart. (7 marks)

OR

18.a Explain ggplot() with and example. (7 marks)

18.b Describe how categorical data is visualized using R. (7 marks)

19.a Illustrate model fitting in simple linear model. (7 marks)

19.b Explain different types of regression. (7 marks)

OR

20.a Describe the unusual observations in the regression model. (7 marks)

20.b Explain corrective measures of unusual observations in regression modelling. (7 marks)

TEACHING PLAN

No	Contents	No of Lecture Hours (35 Hours)
	Module -1 (Introduction to R)	**(8 hours)**
1.1	The R Environment- Command Line Interface and Batch processing, R Packages	1 hour
1.2	Variables, Data Types	1 hour
1.3	Vectors- vector operations and factor vectors	1 hour
1.4	List- List operations, Data Frames	1 hour
1.5	Matrices and arrays	1 hour
1.6	Control Statements- If and else, switch, if else	1 hour
1.7	Loops- For loops, While loops, Controlling loops	1 hour
1.8	Functions- Function as arguments, Named arguments	1 hour
	Module -2(Reading and writing data)	**(8 hours)**
2.1	Importing data from Text files and other software, Exporting data	1 hour
2.2	Importing data from databases- Database Connection packages	1 hour
2.3	Missing Data-NA, NULL	1 hour
2.4	Combining data sets, Transformations	1 hour
2.5	Binning Data, Subsets, summarizing functions	1 hour
2.6	Data Cleaning	1 hour
2.7	Finding and removing Duplicate	1 hour
2.8	Sorting	1 hour
	Module -3 (Statistics with R)	**(6 hours)**
3.1	Analyzing Data	1 hour
3.2	Summary statistics	1 hour
3.3	Statistical Tests- Continuous Data, Discrete Data, Power tests	1 hour
3.4	Common distributions- type arguments	1 hour
3.5	Probability distributions	1 hour
3.6	Normal distributions	1 hour
	Module -4(Data Visualization)	**(6 hours)**
4.1	R Graphics- Overview	1 hour
4.2	Customizing Charts	1 hour
4.3	Graphical parameters, Basic Graphics functions	1 hour
4.4	Lattice Graphics - Lattice functions	1 hour
4.5	Customizing Lattice Graphics	1 hour
4.6	ggplot	1 hour
	Module - 5 (Regression Models)	**(7 hours)**

5.1	Building linear models - model fitting	1 hour
5.2	Predict values using models, Analyzing the fit, Refining the model	1 hour
5.3	Regression- types of regression	1 hour
5.4	Unusual observations and corrective measures	1 hour
5.5	Comparison of models	1 hour
5.6	Generalized linear models -Logistic Regression, Poisson Regression	1 hour
5.7	Nonlinear least squares	1 hour

AMT372	MACHINE LEARNING MODELS AND STORAGE MANAGEMENT	Category	L	T	P	Credit	Year of Introduction
		PEC	2	1	0	3	2020

Preamble: This course enables the learners to understand the basic machine learning models and different storage concepts. The course covers the standard and most popular supervised learning algorithms, storage technology, storage architecture, network storage system and securing and managing storage infrastructures. This course helps the students to choose the appropriate storage infrastructure for typical real world applications.

Prerequisite: Nil

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

CO1	Illustrate the concepts of machine learning techniques and models**(Cognitive Knowledge Level: Apply)**
CO2	Demonstrate various storage management technologies (**Cognitive Knowledge Level: Apply)**
CO3	Explain Storage Systems Architecture and interaction of file systems (**Cognitive Knowledge Level: Understand)**
CO4	Explain the different Network storage protocols **(Cognitive Knowledge Level: Understand)**
CO5	Illustrate the concepts of management metric and standards**(Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

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

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

Assessment Pattern

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

Mark Distribution

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

Continuous Internal Evaluation Pattern:

Attendance **10 marks**

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

Continuous Assessment Assignment **15 marks**

Internal Examination Pattern

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

marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly completed module), each with 7 marks. Out of the 7 questions, a student should answer any 5.

End Semester Examination Pattern:

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

SYLLABUS

Module – 1 (MACHINE LEARNING MODELS)

Introduction to Machine Learning, Examples of Machine Learning applications, Linear Regression: single & multiple variables, Classification: Logistic Regression - Decision Trees, Overfitting & Underfitting, Bias -Variance trade-off, Support Vector Machines, Canonical Cases for Conditional Independence-Naive Bayes' Classifier.

Module - 2(STORAGE TECHNOLOGY)

Information Storage-Data, Bigdata, Information, evolution of storage Architecture. Data Centre Infrastructure-Core elements, characteristics, Virtualization and Cloud Computing, Disk drive components, Physical disk structure, Zone Bit recording, Logical block addressing, Disk drive Performance, Direct Attached Storage, Storage design based on application requirements disk performance

Module- 3(STORAGE SYSTEM ARCHITECTURE)

RAID, Implementation methods, RAID -Techniques-Striping, Mirroring, Parity. RAID Levels, RAID impact on disk performance. Components of an Intelligent Storage System-Front end, Cache, Back end, Storage provisioning-traditional vs virtual. Types of Intelligent storage systems

Backup and Archive- Backup Purpose, Backup Granularity, Backup methods , Backup architectures, Backup topologies

Module - 4 (NETWORK STORAGE SYSTEM)

Fibre Channel Storage Area Networks- SAN and Its Evolution, Components of FC SAN, Fibre Channel Architecture, Fibre Channel Protocol Stack, FC SAN Topologies, Virtualization in SAN, IP SAN and FCoE- iSCSI- Components, FCIP Protocol Stack, Topology, FCoE.

Network-Attached Storage- Benefits of NAS, File Systems and Network File Sharing, Components of NAS, NAS Implementations-Unified NAS, Unified NAS Connectivity, Gateway NAS, Connectivity, NAS File-Sharing Protocols.

Module - 5 (SECURING AND MANAGING STORAGE INFRACTURES)

Information Security Framework, Risk Triad, Storage Security Domains- Securing the Application Access Domain, Securing the Management Access Domain, Securing Backup, Replication, and Archive. Security Implementations in Storage Networking-FC SAN, NAS, IP SAN, Monitoring the Storage Infrastructure, Storage Infrastructure Management Activities, Storage Infrastructure Management Challenges, Information Lifecycle Management, Storage Tiering.

Text Books

1. Introduction to machine learning, Second Edition, EthemAlpayd The MIT Press Cambridge, Massachusetts London, England
2. Information Storage and Management: Storing, Managing, and Protecting Digital Information in Classic, Virtualized, and Cloud Environments, Somasundaram, Gnanasundaram, Alok Shrivastava Editor: EMC Education Services, Wiley, 2012 .

Reference Books

1. Information Storage and Management: Storing, Managing, and Protecting Digital Information, Antonio Cantiago, Wiley, 2009
2. Storage Area Network Essentials: A Complete Guide To Understanding And Implementing Sans, Richard Barker, Paul Massiglia, 2008
3. Storage Networks Explained: Basics and Application of Fibre Channel SAN, NAS, ISCSI, InfiniBand and FCoE, Ulf Troppens and Rainer Erkens, Wiley, 2009

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Compare different machine learning paradigms with suitable examples.
2. Distinguish between overfitting and underfitting. How it can affect model generalization?

Course Outcome 2 (CO2):

1. What is structured and unstructured data? Research the challenges of storing and managing unstructured data.
2. Discuss the benefits of information-centric storage architecture over server-centric storage architecture.

Course Outcome 3 (CO3):

1. What is zoning? Discuss a scenario: a. Where WWN zoning is preferred over port zoning. b. Where port zoning is preferred over WWN zoning.
2. Describe the process of assigning an FC address to a node when logging on to the network for the first time.
3. Seventeen switches, with 16 ports each, are connected in a full mesh topology. How many ports are available for host and storage connectivity?

Course Outcome 1 (CO4):

1. SAN is configured for a backup–to-disk environment, and the storage configuration has additional capacity available. Can you have a NAS gateway configuration use this SAN-attached storage? Discuss the implications ofsharing the backup-to-disk SAN environment with NAS.
2. Compared to a standard IP packet, what percentage of reduction can be realized in protocol overhead in an iSCSI, configured to use jumbo frames with an MTU value of 9,000 bytes?

Course Outcome 5 (CO5):

1. Describe Storage Management strategies for any two real world application scenarios (Storage Allocation to a New Server/Host, File System Space Management)

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES : 4**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: AMT372

Course Name: Machine Learning Models and Storage Management

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

PART A

Answer All Questions. Each Question Carries 3 Marks

1.	Explain the significance of Naive assumption in Bayesian classifier	**3**
2.	Compare Classification with regression with an example	**3**
3.	What are the advantages of a virtualized data center over a classic data center?	**3**
4.	Which components constitute the disk service time? Which component contributes the largest percentage of the disk service time in a random I/O operation?	**3**
5.	What is meant by intelligent storage system.	**3**
6.	Why is RAID 1 not a substitute for a backup?	**3**
7.	Compare the topologies of FC-SAN, NAS, IP-SAN.	**3**
8.	What are the Factors affecting NAS performance.	**3**
9.	List the different security goals.	**3**
10.	How does the use of jumbo frames affect the NAS performance?	**3**

(10x3=30)

Part B

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

11. (a) For the following set of training samples, find which attribute can be chosen as the root for decision tree classification **(8)**

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

(b) Explain the working principles of SVM classifiers. **(6)**

OR

12. (a) What is overfitting? Explain Bias -Variance trade off. **(7)**

(b) Use the following data to construct a linear regression model for the auto insurance premium as a function of number of years the vehicle used. **(7)**

Years used	1	3	5	8	10	12
Insurance Premium	9000	7000	6000	5000	4000	3000

13. (a) The average I/O size of an application is 64 KB. The following specifications are available from the disk manufacturer: average seek time = 5 ms, 7,200 RPM, and transfer rate = 40 MB/s. Determine the maximum IOPS that could be performed with this disk for the application. Using this case as an example, explain the relationship between disk utilization and IOPS **(7)**

(b) Illustrate any three Disk Drive Components. **(7)**

OR

14. (a) Define the following terms **(8)**

(i)Disk Service Time
(ii)Seek Time
(iii)Rotational Latency
(iv)Data Transfer Rate

(b) List the benefits and limitations of Direct Attached Storage **(6)**

15. (a) Explain the terms :Striping, Mirroring, Parity **(6)**

(b) Describe the Components of an Intelligent Storage System **(8)**

OR

16. (a) Explain the process of data recovery in case of a drive failure in RAID 5. What are the benefits of using RAID 3 in a backup application? **(7)**

(b) Explain the Array caching properties and algorithms. **(7)**

17. (a) Illustrate the NAS File-Sharing Protocols. **(10)**

(b) Explain Fibre Channel Architecture and Protocol Stack. **(4)**

OR

18. (a) Describe the Benefits of CAS? **(8)**

(b) Explain the Components of IP-SAN? **(6)**

19. (a) Explain how security is provided in application access domain and management access domain. **(10)**

(b) List out the challenges in storage Infrastructure management **(4)**

OR

20. (a) Describe the secure user access in NAS environment **(6)**

(b) Discuss different aspects of monitoring the storage infrastructure **(8)**

Teaching Plan

No	Contents	No. of Lecture Hours (37hrs)
	Module – 1 (FUNDAMENTALS) (7 hours)	
1.1	Introduction to Machine Learning, Examples of Machine Learning applications	1 hour
1.2	Linear Regression: single & multiple variables,	1 hour
1.3	Classification: Logistic Regression	1 hour
1.4	Decision Trees	1 hour
1.5	Overfitting & Underfitting, Bias Variance Trade-off	1 hour
1.6	Support Vector Machines	1 hour
1.7	Canonical Cases for Conditional Independence-Naive Bayes' Classifier.	1 hour
	Module - 2(STORAGE TECHNOLOGY) (5 hours)	
2.1	Information Storage-Data, Bigdata, Information, evolution of storage Architecture	1 hour
2.2	Data Centre Infrastructure-Core elements, characteristics, Virtualization and Cloud Computing	1 hour
2.3	Disk drive components, Physical disk structure, Zone Bit recording, Logical block addressing	1 hour
2.4	Disk drive Performance, Direct Attached Storage	1 hour
2.5	Storage design based on application requirements disk performance	1 hour
	Module - 3 (STORAGE SYSTEM ARCHITECTURE) (8 hours)	

3.1	RAID, Implementation methods, RAID -Techniques-Striping, Mirroring, Parity.	
3.2	RAID Levels, RAID impact on disk performance	
3.3	Components of an Intelligent Storage System-Front end, Cache, Back end,	
3.4	Storage provisioning-traditional vs virtual.	
3.5	Types of Intelligent storage systems	
3.6	Backup and Archive- Backup Purpose	
3.7	Backup Granularity, Backup methods , Backup architectures	
3.8	Backup topologies	
Module - 4 (NETWORK STORAGE SYSTEM) (10 hours)		
4.1	Fibre Channel Storage Area Networks- SAN and Its Evolution, Components of FC SAN	1 hour
4.2	Fibre Channel Architecture, Fibre Channel Protocol Stack	1 hour
4.3	Zoning	1 hour
4.4	FC SAN Topologies, Virtualization in SAN	1 hour
4.5	IP SAN- FCoE and iSCSI, Components, topologies	1 hour
4.6	iSCSI Protocol stack	1 hour
4.7	FCoE, Components of FCoE	1 hour
4.8	Network-Attached Storage- Benefits of NAS, File Systems and Network File Sharing,	
4.9	Components of NAS, NAS Implementations-Unified NAS, Unified NAS Connectivity, Gateway NAS, Connectivity,	
4.10	NAS File-Sharing Protocols	
Module - 5 (MANAGING AND MONITORING) (7 hours)		
5.1	Managing & Monitoring: Management philosophies	1 hour
5.2	Industry management standards (SNMP, SMI-S, CIM)	1 hour
5.3	Standard framework applications, Key management metrics (thresholds,	1 hour

	availability, capacity, security, performance)	
5.4	Standard framework applications, Key management metrics (thresholds, availability, capacity, security, performance)	1 hour
5.5	Provisioning & configuration change planning	1 hour
5.6	Problem reporting	1 hour
5.7	prioritization and handling techniques, Management tools overview	1 hour

CST 332	FOUNDATIONS OF SECURITY IN COMPUTING	Category	L	T	P	Credit	Year Of Introduction
		PEC	2	1	0	3	2019

Preamble: The purpose of this course is to create awareness among learners about the fundamentals of security and number theory. This course covers Integer & Modular Arithmetic, Primes & Congruences, Discrete Logarithms & Elliptic Curve Arithmetic and an overview of computer security. The concepts covered in this course enable the learners in effective use of cryptographic algorithms and to identify the security threats in computing.

Prerequisite: A sound knowledge in Mathematics, Discrete Computational Structures, Operating Systems and Database Systems.

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

CO1	Illustrate the operations and properties of algebraic structures, integer arithmetic and modular arithmetic. **(Cognitive Knowledge Level: Understand)**
CO2	Use the concepts of prime numbers and factorization for ensuring security in computing systems **(Cognitive Knowledge Level: Apply)**
CO3	Illustrate the concepts of Linear Congruence, Primitive Roots, Discrete Logarithms and Elliptic Curve Arithmetic **(Cognitive Knowledge Level: Apply)**
CO4	Summarize the threats and attacks related to computer and program security **(Cognitive Knowledge Level: Understand)**
CO5	Outline the key aspects of operating system and database security **(Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

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

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

Assessment Pattern

Bloom's Category	Test 1 (%)	Test 2 (%)	End Semester Examination (%)
Remember	**30**	**30**	**30**
Understand	**30**	**30**	**30**
Apply	**40**	**40**	**40**
Analyse			

Evaluate			
Create			

Mark Distribution

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

Continuous Internal Evaluation Pattern:

Attendance : **10 marks**

Continuous Assessment Test : **25 marks**

Continuous Assessment Assignment : **15 marks**

Internal Examination Pattern:

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

End Semester Examination Pattern:

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

Syllabus

Module-1 (Modular Arithmetic)

Integer arithmetic - Integer division, Divisibility, Greatest Common Divisor (GCD), Euclid's algorithm for GCD, Extended Euclid’s algorithm, Linear Diophantine Equations. Modular

arithmetic - Operations, Properties. Algebraic structures - Groups, Rings, Fields, Finite fields, GF(p), GF (2^n).

Module-2 (Prime Numbers and Factorization)

Prime numbers - Prime numbers and prime-power factorization, Fermat and Mersenne primes, Fermat's theorem, Applications, Euler's theorem, Euler's totient function, Applications. Primality testing – Deterministic algorithms and Probabilistic algorithms. Factorization - Fermat's factorization, Pollard p-1 method.

Module-3 (Linear Congruence, Primitive Roots and Elliptic Curve Arithmetic)

Linear congruence - Simultaneous linear congruence, Chinese Remainder Theorem (CRT). Congruence with a prime - Power modulus, Arithmetic modulo p, Pseudoprimes and Carmichael numbers, Solving congruence modulo prime powers. Primitive roots - Existence of primitive roots for primes, Discrete logarithms. Elliptic curve arithmetic – Prime curves, Binary curves, Addition of two points, Multiplication of a point by a constant.

Module-4 (Computer and Program Security)

Introduction to computer security – Threats, Vulnerabilities, Controls. Browser attack types, Web attacks targeting users, Email attack types. Introduction to program security - Non-malicious programming oversights, Malware.

Module-5 (Operating System and Database Security)

Operating system security – Security in operating system, Security in design of operating system. Database security – Security requirements of databases, Reliability and integrity, Database disclosure.

Text Books

1. Behrouz A Forouzan, Cryptography and Network Security, 3/e, Tata McGraw-Hill.
2. Charles P Pfleeger, Shari Lawrence Pfleeger, Jonathan Margulies, Security in Computing, 5/e, Prentice Hall.
3. G.A. Jones & J.M. Jones, Elementary Number Theory, Springer UTM, 2007

References

1. William Stallings, Cryptography and Network Security Principles and Practices, 4/e, Pearson Ed.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Find the n- bit word that is represented by the polynomial $x^2 + 1$ in $GF(2^5)$.
2. Solve the linear Diophantine equation 21x + 14y=35.

Course Outcome 2 (CO2):

1. Prove that a Carmichael number cannot be the product of two distinct primes.
2. Use the Pollard p-1 method to find a factor of 57247159 with the bound B=8.

Course Outcome 3 (CO3):

1. Find an integer that has a remainder of 3 when divided by 7 and 13, but is divisible by 12.
2. In the elliptic curve E(1,2) over the field GF(11), find the equation of the curve and all the points on the curve.

Course Outcome 4 (CO4):

1. List three controls that could be applied to detect or prevent off-by-one errors.
2. How does fake email messages act as spam?

Course Outcome 5 (CO5):

1. Discuss the importance of auditability and access control in database security.
2. Explain the various factors which can make data sensitive.

Model Question Paper

QP CODE: **PAGES: ___**

Reg No:_____
Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: CST 332
Course Name : FOUNDATIONS OF SECURITY IN COMPUTING

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

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

1. List the four properties of divisibility with examples.
2. Find gcd (401,700) using Euclid's algorithm.
3. Use Fermat's Little theorem to show that 91 is not a prime.
4. If m is relatively prime to n , show that $\Phi(mn) = \Phi(m)\,\Phi(n)$.
5. Solve the congruence relation $103x \equiv 57 \pmod{211}$.
6. Find a solution for the congruence $3x \equiv 5 \bmod 7^3$
7. What are the problems created by an off-by-one error?
8. How does a clickjacking attack succeed?
9. Explain the significance of correctness and completeness in the design of operating systems.
10. How does the two-phase update technique help the database manager in handling failures?

(10x3=30)

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

11. (a) For the group G = < Z_6*, x>, prove that it is an Abelian group. Also show the result of 5 x 1 and 1 ÷ 5. **(6)**

(b) Find a particular and the general solution to the following linear Diophantine equations. **(8)**
i) 19 x+13y = 20 ii) 40 x + 16 y =88

OR

12. (a) Describe the properties of modular arithmetic and modulo operator. **(6)**

(b) Using Extended Euclidean algorithm, find the multiplicative inverse of (i) 131 in Z_{180} and (ii) 23 in Z_{100}. **(8)**

13. (a) State and prove Fermat's theorem. **(6)**

(b) Explain Fermat's factorization method and use it to factor 809009. **(8)**

OR

14. (a) Define Euler's totient function. Prove that, Ø(pq)=(p-1)(q-1) where p and q are prime numbers. **(7)**

(b) Define Fermat primes. Show that any two distinct Fermat numbers are relatively prime. **(7)**

15. (a) Using Chinese Remainder Theorem, solve the system of congruence, x ≡2(mod 3), x ≡3(mod 5), x ≡2(mod 7). **(7)**

(b) Define Carmichael number and show that a Carmichael number must be the product of at least three distinct primes. **(7)**

OR

16. (a) For the group G = <Z_{19*}, x>, find the primitive roots in the group. **(6)**

(b) Consider the elliptic curve $y^2 = x^3 + x + 1$ defined over Z_{23}. If P = (3, 10) and Q = (9,7) are two points on the elliptic curve, find 2P and P + Q. **(8)**

17. (a) Distinguish the terms vulnerability, threat and control. **(4)**

(b) With the help of suitable examples, explain the security problems created by incomplete mediation and time-of-check to time-of use. **(10)**

OR

18. (a) Differentiate between man-in-the-browser attack and page-in-the-middle attack. **(4)**

(b) Explain the four aspects of malicious code infection. **(10)**

19. (a) List any six computer security related functions addressed by operating systems. **(6)**

(b) How does a kernelized design support in enforcing security mechanisms? **(8)**

OR

20. (a) Explain any four security requirements of databases. **(4)**

(b) How can database disclosure be prevented? With the help of suitable examples, explain any six types of disclosure. **(10)**

Teaching Plan

No	Contents	No.of Lecture Hrs
	Module-1 (Modular Arithmetic) (6 hrs)	
1.1	Integer arithmetic, Integer division, Divisibility, Greatest Common Divisor (GCD)	1
1.2	Euclid's algorithm for GCD, Extended Euclid's algorithm	1
1.3	Linear Diophantine Equations	1
1.4	Modular arithmetic operations, Properties of modular arithmetic	1
1.5	Groups, Rings and Fields	1
1.6	Finite fields – GF(p), GF(2^n)	1
	Module-2 (Prime Numbers and Factorization) (7 hrs)	
2.1	Prime numbers and prime-power factorization	1
2.2	Fermat and Mersenne primes	1
2.3	Fermat's theorem, Applications – Exponentiation, Multiplicative inverse	1
2.4	Euler's theorem, Euler's totient function, Applications	1
2.5	Primality testing – Deterministic algorithms – Divisibility algorithm	1

2.6	Primality testing – Probabilistic algorithms-Fermat test, Square root test, Miller - Rabin test	1
2.7	Factorization - Fermat's factorization, Pollard p-1 method	1
Module-3 (Linear Congruence, Primitive Roots and Elliptic Curve Arithmetic) (7 hrs)		
3.1	Linear congruence, Simultaneous linear congruence	1
3.2	Chinese Remainder Theorem (CRT)	1
3.3	Congruence with a Prime-Power Modulus, Arithmetic modulo p	1
3.4	Pseudo-primes and Carmichael numbers	1
3.5	Solving congruence modulo prime powers	1
3.6	Primitive roots, Existence of primitive roots for primes, Discrete logarithms	1
3.7	Elliptic curve arithmetic – Prime curves, Binary curves, Addition of two points, Multiplication of a point by a constant	1
Module-4 (Computer and Program Security) (7 hrs) (Text book2: Chapters 1, 3, 4)		
4.1	Threats, Vulnerabilities, Controls	1
4.2	Browser attack types	1
4.3	Web attacks targeting users	1
4.4	Email attack types	1
4.5	Non-malicious programming oversights (Lecture 1)	1
4.6	Non-malicious programming oversights (Lecture 2)	1
4.7	Malware – Four aspects of infection	1
Module-5 (Operating System and Database Security) (8 hrs)(Text book2: Chapters 5, 7)		
5.1	Security in operating system (Lecture 1)	1
5.2	Security in operating system (Lecture 2)	1
5.3	Security in design of operating system (Lecture 1)	1

5.4	Security in design of operating system (Lecture 2)	1
5.5	Security requirements of databases	1
5.6	Reliability & integrity	1
5.7	Database disclosure (Lecture 1)	1
5.8	Database disclosure (Lecture 2)	1

HUT 300	Industrial Economics & Foreign Trade	Category	L	T	P	CREDIT
		HSMC	3	0	0	3

Preamble: To equip the students to take industrial decisions and to create awareness of economic environment.

Prerequisite: Nil

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

CO1	Explain the problem of scarcity of resources and consumer behaviour, and to evaluate the impact of government policies on the general economic welfare. (Cognitive knowledge level: **Understand**)
CO2	Take appropriate decisions regarding volume of output and to evaluate the social cost of production. (Cognitive knowledge level: **Apply**)
CO3	Determine the functional requirement of a firm under various competitive conditions. (Cognitive knowledge level: **Analyse**)
CO4	Examine the overall performance of the economy, and the regulation of economic fluctuations and its impact on various sections in the society. (Cognitive knowledge level: **Analyse**)
CO5	Determine the impact of changes in global economic policies on the business opportunities of a firm. (Cognitive knowledge level: **Analyse**)

Mapping of course outcomes with program outcomes

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

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

Assessment Pattern

Bloom's Category	**Continuous Assessment Tests**		**End Semester Examination Marks**
	Test 1 (Marks)	**Test 2 (Marks)**	
Remember	**15**	**15**	**30**
Understand	**20**	**20**	**40**
Apply	**15**	**15**	**30**

Mark Distribution

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

Continuous Internal Evaluation Pattern:

Attendance	: 10 marks
Continuous Assessment - Test (2 numbers)	: 25 marks
Continuous Assessment - Assignment	: 15 marks

Internal Examination Pattern:

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

End Semester Examination Pattern:

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

Part A	: 30 marks
Part B	: 70 marks

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

SYLLABUS

HUT 300 Industrial Economics & Foreign Trade

Module 1 (Basic Concepts and Demand and Supply Analysis)

Scarcity and choice - Basic economic problems- PPC – Firms and its objectives – types of firms – Utility – Law of diminishing marginal utility – Demand and its determinants – law of demand – elasticity of demand – measurement of elasticity and its applications – Supply, law of supply and determinants of supply – Equilibrium – Changes in demand and supply and its effects – Consumer surplus and producer surplus (Concepts) – Taxation and deadweight loss.

Module 2 (Production and cost)

Production function – law of variable proportion – economies of scale – internal and external economies – Isoquants, isocost line and producer's equilibrium – Expansion path – Technical progress and its implications – Cobb-Douglas production function - Cost concepts – Social cost: private cost and external cost – Explicit and implicit cost – sunk cost - Short run cost curves - long run cost curves – Revenue (concepts) – Shutdown point – Break-even point.

Module 3 (Market Structure)

Perfect and imperfect competition – monopoly, regulation of monopoly, monopolistic completion (features and equilibrium of a firm) – oligopoly – Kinked demand curve – Collusive oligopoly (meaning) – Non-price competition – Product pricing – Cost plus pricing – Target return pricing – Penetration pricing – Predatory pricing – Going rate pricing – Price skimming.

Module 4 (Macroeconomic concepts)

Circular flow of economic activities – Stock and flow – Final goods and intermediate goods - Gross Domestic Product - National Income – Three sectors of an economy- Methods of measuring national income – Inflation- causes and effects – Measures to control inflation- Monetary and fiscal policies – Business financing- Bonds and shares -Money market and Capital market – Stock market – Demat account and Trading account - SENSEX and NIFTY.

Module 5 (International Trade)

Advantages and disadvantages of international trade - Absolute and Comparative advantage theory - Heckscher - Ohlin theory - Balance of payments – Components – Balance of Payments

deficit and devaluation – Trade policy – Free trade versus protection – Tariff and non-tariff barriers.

Reference Materials

1. Gregory N Mankiw, 'Principles of Micro Economics', Cengage Publications
2. Gregory N Mankiw, 'Principles of Macro Economics', Cengage Publications
3. Dwivedi D N, 'Macro Economics', Tata McGraw Hill, New Delhi.
4. Mithani D M, 'Managerial Economics', Himalaya Publishing House, Mumbai.
5. Francis Cherunilam, 'International Economics', McGraw Hill, New Delhi.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Why does the problem of choice arise?
2. What are the central problems?
3. How do we solve the basic economic problems?
4. What is the relation between price and demand?
5. Explain deadweight loss due to the imposition of a tax.

Course Outcome 2 (CO2):

1. What is shutdown point?
2. What do you mean by producer equilibrium?
3. Explain break-even point;
4. Suppose a chemical factory is functioning in a residential area. What are the external costs?

Course Outcome 3 (CO3):

1. Explain the equilibrium of a firm under monopolistic competition.
2. Why is a monopolist called price maker?
3. What are the methods of non-price competition under oligopoly?

4. What is collusive oligopoly?

Course Outcome 4 (CO4):

1. What is the significance of national income estimation?
2. How is GDP estimated?
3. What are the measures to control inflation?
4. How does inflation affect fixed income group and wage earners?

Course Outcome 5 (CO5):

1. What is devaluation?
2. Suppose a foreign country imposes a tariff on Indian goods. How does it affect India's exports?
3. What is free trade?
4. What are the arguments in favour of protection?

Model Question paper

QP CODE: **PAGES:3**

Reg No:_____ **Name :_____**

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

Course Code: HUT 300

Course Name: Industrial Economics & Foreign Trade

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

PART A

Answer all Questions. Each question carries 3 Marks

1. Why does an economic problem arise?

2. What should be the percentage change in price of a product if the sale is to be increased by 50 percent and its price elasticity of demand is 2?

3. In the production function Q= $2L^{1/2}K^{1/2}$ if L=36 how many units of capital are needed to produce 60 units of output?

4. Suppose in the short run AVC 4. Suppose in the short run AVC<P<AC. Will this firm produce or shut down? Give reason.

5. What is predatory pricing?

6. What do you mean by non- price competition under oligopoly?

7. What are the important economic activities under primary sector?

8. Distinguish between a bond and share?

9. What are the major components of balance of payments?

10. What is devaluation? **(10 x 3 = 30 marks)**

PART B

(Answer one full question from each module, each question carries 14 marks)

MODULE I

11. a) Prepare a utility schedule showing units of consumption, total utility and marginal utility, and explain the law of diminishing marginal utility. Point out any three limitations of the law.

    b) How is elasticity of demand measured according to the percentage method? How is the measurement of elasticity of demand useful for the government?

Or

12. a) Explain the concepts consumer surplus and producer surplus.

    b) Suppose the government imposes a tax on a commodity where the tax burden met by the consumers. Draw a diagram and explain dead weight loss. Mark consumer surplus, producer surplus, tax revenue and dead weight loss in the diagram.

MODULE II

13. a) What are the advantages of large-scale production?

    b) Explain Producer equilibrium with the help of isoquants and isocost line. What is expansion path?

Or

14. a) Explain break-even analysis with the help of a diagram.

    b) Suppose the monthly fixed cost of a firm is Rs. 40000 and its monthly total variable cost is Rs. 60000.

    i. If the monthly sales is Rs. 120000 estimate contribution and break-even sales.
    ii. If the firm wants to get a monthly profit of Rs.40000, what should be the sales?

    c) The total cost function of a firm is given as $TC=100+50Q - 11Q^2+Q^3$. Find marginal cost when output equals 5 units.

MODULE III

15. a) What are the features of monopolistic competition?

b) Explain the equilibrium of a firm earning supernormal profit under monopolistic competition.

Or

16. a) Make comparison between perfect competition and monopoly.

b) Explain price rigidity under oligopoly with the help of a kinked demand curve.

MODULE IV

17. a) How is national income estimated under product method and expenditure method?

b) Estimate GDPmp, GNPmp and National income

Private consumption expenditure	= 2000 (in 000 cores)
Government Consumption	= 500
NFIA	= -(300)
Investment	= 800
Net=exports	=700
Depreciation	= 400
Net-indirect tax	= 300

Or

18. a) What are the monetary and fiscal policy measures to control inflation?

b) What is SENSEX?

MODULE V

19. a) What are the advantages of disadvantages of foreign trade?

b) Explain the comparative cost advantage.

Or

20. a) What are the arguments in favour protection?

b) Examine the tariff and non-tariff barriers to international trade.

(5 × 14 = 70 marks)

Teaching Plan

Module 1 (Basic concepts and Demand and Supply Analysis)		**7 Hours**
1.1	Scarcity and choice – Basic economic problems - PPC	1 Hour
1.2	Firms and its objectives – types of firms	1 Hour
1.3	Utility – Law of diminishing marginal utility – Demand – law of demand	1 Hour
1.4	Measurement of elasticity and its applications	1 Hour
1.5	Supply, law of supply and determinants of supply	1 Hour
1.6	Equilibrium – changes in demand and supply and its effects	1 Hour
1.7	Consumer surplus and producer surplus (Concepts) – Taxation and deadweight loss.	1 Hour
Module 2 (Production and cost)		**7 Hours**
2.1	Productions function – law of variable proportion	1 Hour
2.2	Economies of scale – internal and external economies	1 Hour
2.3	producers equilibrium – Expansion path	1 Hour
2.4	Technical progress and its implications – cob Douglas Production function	1 Hour
2.5	Cost concepts – social cost: private cost and external cost – Explicit and implicit cost – sunk cost	1 Hour
2.6	Short run cost curves & Long run cost curves	1 Hour
2.7	Revenue (concepts) – shutdown point – Break-even point.	1 Hour
Module 3 (Market Structure)		**6 hours**
3.1	Equilibrium of a firm, MC – MR approach and TC – TR approach	1 Hour
3.2	Perfect competition & Imperfect competition	1 Hour
3.3	Monopoly – Regulation of monopoly – Monopolistic competition	1 Hour
3.4	Oligopoly – kinked demand curve	1 Hour
3.5	Collusive oligopoly (meaning) – Non price competition	1 Hour
3.6	Cost plus pricing – Target return pricing – Penetration, Predatory pricing – Going rate pricing – price skimming	1 Hour

Module 4 (Macroeconomic concepts)		7 Hours
4.1	Circular flow of economic activities	1 Hour
4.2	Stock and flow – Final goods and intermediate goods – Gross Domestic Product - National income – Three sectors of an economy	1 Hour
4.3	Methods of measuring national income	1 Hour
4.4	Inflation – Demand pull and cost push – Causes and effects	1 Hour
4.5	Measures to control inflation – Monetary and fiscal policies	1 Hour
4.6	Business financing – Bonds and shares – Money market and capital market	1 Hour
4.7	Stock market – Demat account and Trading account – SENSEX and NIFTY	1 Hour
Module 5 (International Trade)		**8 Hours**
5.1	Advantages and disadvantages of international trade	1 Hour
5.2	Absolute and comparative advantage theory	2 Hour
5.3	Heckscher – Ohlin theory	1 Hour
5.4	Balance of payments - components	1 Hour
5.5	Balance of payments deficit and devaluation	1 Hour
5.6	Trade policy – Free trade versus protection	1 Hour
5.7	Tariff and non tariff barriers.	1 Hour

ADT308	COMPREHENSIVE COURSE WORK	Category	L	T	P	Credit	Year of Introduction
		PCC	1	0	0	1	2019

Preamble: The objective of this Course work is to ensure the comprehensive knowledge of each studentinthemostfundamentalcorecoursesinthecurriculum.Five core courses credited from semesters 3, 4 and 5 are chosen for the detailed study in this course work. This course helps the learner to become competent in cracking GATE, placement tests and other competitive examinations.

Prerequisite:

1. **Introduction to Machine Learning**
2. **Data Structures**
3. **Operating Systems**
4. **Database Management Systems**
5. **Foundation of Data Science**

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

CO1:	Comprehend the concepts in machine learning (Cognitive Knowledge Level: **Understand)**
CO2:	Comprehend the concepts and applications of data structures (Cognitive Knowledge Level: **Understand)**
CO3 :	Comprehend the concepts, functions and algorithms in Operating System (Cognitive Knowledge Level: **Understand)**)
CO4:	Comprehend the fundamental principles of database design and manipulation (Cognitive Knowledge Level: **Understand**)
CO5:	Comprehend the basic concepts of data science (Cognitive Knowledge Level: **Understand)**

Mapping of course outcomes with program outcomes

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

Assessment Pattern

Bloom's Category	End Semester Examination
Remember	**10**
Understand	**20**
Apply	**20**
Analyse	
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
50	**0**	**50**	**1 hour**

End Semester Examination Pattern: Objective Questions with multiple choice, a maximum of four options. Question paper include fifty questions of one mark each, distributed equally from all the five identified courses.

SYLLABUS

Full Syllabus of all five selected Courses.

1. **Introduction to Machine Learning**
2. **Data Structures**
3. **Operating Systems**
4. **Database Management Systems**
5. **Foundation of Data Science**

Course Contents and Lecture Schedule

No	Topic	No. of Lectures
1	**INTRODUCTION TO MACHINE LEARNING**	
1.1	Mock Test on Module 1, Module 2 and Module 3	1 hour
1.2	Mock Test on Module 4 and Module 5	1 hour
1.3	Feedback and Remedial class	1 hour
2	**DATA STRUCTURES**	
2.1	Mock Test on Module 1, Module 2 and Module 3	1 hour
2.2	Mock Test on Module 4 and Module 5	1 hour
2.3	Feedback and Remedial class	1 hour
3	**OPERATING SYSTEMS**	
3.1	Mock Test on Module 1 and Module 2	1 hour
3.2	Mock Test on Module 3, Module 4 and Module 5	1 hour
3.3	Feedback and Remedial class	1 hour
4	**DATABASE MANAGEMENT SYSTEMS**	
4.1	Mock Test on Module 1, Module 2 and Module 3	1 hour
4.2	Mock Test on Module 4 and Module 5	1 hour

4.3	Feedback and Remedial class	1 hour
5	**FOUNDATIONS OF DATA SCIENCE**	
5.1	Mock Test on Module 1, Module 2 and Module 3	1 hour
5.2	Mock Test on Module 4 and Module 5	1 hour
5.3	Feedback and Remedial class	1 hour

Model Question Paper

QP CODE:

Reg No: _____

Name: _____ **PAGES :7**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

SIXTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: ADT308

Course Name: Comprehensive Course Work

Max. Marks: 50 **Duration: 1 Hour**

Objective type questions with multiple choices. Mark one correct answer for each question. Each Question Carries 1 Mark

1. Application of machine learning methods to large databases is called

   (A) Data Mining (B) Artificial Intelligence

   (C) Big Data Computing (D) Internet of Things

2. If machine learning model output involves target variable, then that model is called as

   (A) Descriptive Model (B) Predictive Model

   (C) Reinforcement Learning (D) All of the above

3. In what type of learning labelled training data is used

   (A) Unsupervised Learning (B) Supervised Learning

   (C) Reinforcement Learning (D) Active Learning

4. In following type of feature selection method we start with empty feature set

(A) Forward Feature Selection (B) Backward Feature Selection

(C) Both A and B (D) None of the above

5. Which of the following is the best machine learning method?

(A) Scalable (B) Accuracy

(C) Fast (D) All of the above

6. Data used to build a data mining model.

(A) Training data (B) Validation data

(C) Test data (D) Hidden data

7. You are given reviews of few netflix series marked as positive, negative and neutral. Classifying reviews of a new netflix series is an example of

(A) Supervised learning (B) Unsupervised learning

(C) Semisupervised learning (D) Reinforcement learning

8. Following are the types of supervised learning

(A) Classification (B) Regression

(C) subgroup discovery (D) all of the above

9. The output of training process in machine learning is

(A) machine learning model (B) machine learning algorithm

(C) null (D) accuracy

10. PCA is

(A) forward feature selection (B) backward feature selection

(C) feature extraction (D) all of the above

11. Consider the following sequence of operations on an empty stack.
push(22); push(43); pop(); push(55); push(12); s=pop();
Consider the following sequence of operations on an empty queue.
enqueue(32);enqueue(27); dequeue(); enqueue(38); enqueue(12); q=dequeue();
The value of s+q is _____
(A) 44 (B) 54 (C) 39 (D) 70

12. A B-tree of order (degree)5 and of height 3 will have a minimum of ___ keys.
A. 624
B. 249
C. 124
D. 250

13. Construct a binary search tree by inserting 8, 6, 12, 3, 10, 9 one after another. To make the resulting tree as AVL tree which of the following is required?
    (A) One right rotation only
    (B) One left rotation followed by two right rotations
    (C) One left rotation and one right rotation
    (D) The resulting tree itself is AVL

14. In a complete 4-ary tree, every internal node has exactly 4 children or no child. The number of leaves in such a tree with 6 internal nodes is:
    (A) 20 (B) 18 (C) 19 (D) 17

15. Select the postfix expression for the infix expression *a+b-c+d*(e/f)*.
    (A) ab+c-d+e*f/ (B) ab+c-def/*+
    (C) abc-+def/*+ (D) ab+c-def/*+

16. Consider a hash table of size seven, with starting index zero, and a hash function (2x + 5)mod7. Assuming the hash table is initially empty, which of the following is the contents of the table when the sequence 1, 4, 9, 6 is inserted into the table using closed hashing? Note that '_' denotes an empty location in the table.
    (A) 9, _, 1, 6, _, _, 4 (B) 1, _, 6, 9, _, _, 4
    (C) 4, _, 9, 6, _, _, 1 (D) 1, _, 9, 6, _, _, 4

17. **Compute the time complexity of the following function:**

```
void function(int n)
{
        int count = 0;
        for (int i=n/2; i<=n; i++)
                for (int j=1; j<=n; j = j + 2)
                        for (int k=1; k<=n; k = k * 2)
                                count++;}
```

    A. $O(n^2 \log n)$
    B. $O(n \log^2 n)$
    C. $O(n^3)$
    D. $O(n \log n^2)$

18. How many distinct binary search trees can be created out of 6 distinct keys?
    (A) 7 (B) 36 (C) 140 (D) 132

19. Which tree traversal performed on a binary search tree, results in ascending order listing of the keys?
    A. Pre-order
    B. In-order
    C. Post-order
    D. Level-order

20. You are given pointers to first and last nodes of a singly linked list, which of the following operations are dependent on the length of the linked list?
    (A) Delete the first element
    (B) Insert a new element as a first element
    (C) Add a new element at the end of the list
    (D) Delete the last element of the list

21. Suppose a disk has 400 cylinders, numbered from 0 to 399. At some time the disk arm is at cylinder 58, and there is a queue of disk access requests for cylinder 66, 349, 201, 110, 38, 84, 226, 70, 86. If Shortest-Seek Time First (SSTF) is being used for scheduling the disk access, the request for cylinder 86 is serviced after servicing _____ number of requests.
    (A) 1 (B) 2 (C)3 (D)4

22. If frame size is 4KB then a paging system with page table entry of 2 bytes can address _____ bytes of physical memory.
    (A) 2^12 (B) 2^16 (C) 2^18 (D) 2^28

23. Calculate the internal fragmentation if page size is 4KB and process size is 103KB.
    (A) 3KB (B) 4KB (C) 1KB (D) 2KB

24. Which of the following scheduling policy is likely to improve interactiveness?
    (A) FCFS (B) Round Robin
    (C) Shortest Process Next (D) Priority Based Scgeduling

25. Consider the following program

```
        Semaphore X=1, Y=0
Void A ( )                          Void B ( )
{                                   {
        While (1)                           While (1)
        {                                   {
        P(X);                               P(Y);
        Print'1';                           P(X);
        V(Y);                               Print'0';
        }                                   V(X);
}                                           }
                                    }
```

    The possible output of the program:
    (A) Any number of 0's followed by any number of 1's.
    (B) Any number of 1's followed by any number of 0's.
    (C) 0 followed by deadlock
    (D) 1 followed by deadlock

26. In a system using single processor, a new process arrives at the rate of 12 processes per minute and each such process requires 5 seconds of service time. What is the percentage of CPU utilization?

(A) 41.66 (B) 100.00 (C) 240.00 (D) 60.00

27. A system has two processes and three identical resources. Each process needs two resources to proceed. Then
(A) Deadlock is possible (B) Deadlock is not possible
(C) Starvation may be present (D) Thrashing

28. Which of the following is true with regard to Round Robin scheduling technique?
(A) Responds poorly to short process with small time quantum.
(B) Works like SJF for larger time quantum
(C) Does not use a prior knowledge of burst times of processes.
(D) Ensure that the ready queue is always of the same size.

29. Thrashing can be avoided if
(A) the pages, belonging to working set of programs, are in main memory
(B) the speed of CPU is increased
(C) the speed of I/O processor is increased
(D) none of the above

30. The circular wait condition can be prevented by
(A) using thread
(B) defining a linear ordering of resource types
(C) using pipes
(D) all of the above

31. Let E1, E2 and E3 be three entities in an E/R diagram with simple single-valued attributes. R1 and R2 are two relationships between E1 and E2, where R1 is one-to-many, R2 is many-to-many. R3 is another relationship between E2 and E3 which is many-to-many. R1, R2 and R3 do not have any attributes of their own. What is the minimum number of tables required to represent this situation in the relational model?
(A) 3 (B) 4 (C) 5 (D) 6

32. Identify the minimal key for relational scheme R(U, V, W, X, Y, Z) with functional dependencies F = {U → V, V → W, W → X, VX → Z}
(A) UV (B) UW (C) UX (D) UY

33. It is given that: “Every student need to register one course and each course registered by many students”, what is the cardinality of the relation say “Register” from the “Student” entity to the “Course” entity in the ER diagram to implement the given requirement.
(A) M:1 relationship (B) M:N relationship
(C) 1:1 relationship (D) option (B) or(C)

34. Consider the relation branch(branch_name, assets, branch_city)
SELECT DISTINCT T.branch_name FROM branch T, branch S WHERE T.assets>L.assets AND S.branch_city = "TVM" .
Finds the names of
(A) All branches that have greater assets than all branches located in TVM.

(B) All branches that have greater assets than some branch located in TVM.
(C) The branch that has the greatest asset in TVM.
(D) Any branch that has greater asset than any branch located in TVM.

35. Consider the following relation instance, where "A" is primary Key.

A1	A2	A3	A4
1	1	1	Null
5	2	5	1
9	5	13	5
13	13	9	15

Which one of the following can be a foreign key that refers to the same relation?
(A) A2 (B) A3 (C) A4 (D) ALL

36. A relation R(ABC) is having the tuples(1,2,1),(1,2,2),(1,3,1) and (2,3,2). Which of the following functional dependencies holds well?
(A) A → BC (B) AC → B (C) AB → C (D) BC → A

37. Consider a relation R with attributes A, B, C, D and E and functional dependencies A→BC, BC→E, E→DA. What is the highest normal form that the relation satisfies?
(A) BCNF (B) 3 NF (C) 2 NF (D) 1 NF

38. For the given schedule S, find out the conflict equivalent schedule.
S : r1(x); r2(Z) ; r3(X); r1(Z); r2(Y); r3(Y);W1(X); W2(Z); W3(Y); W2(Y)
(A) T1→T2→T3 (B) T2->T1->T3
(C) T3→T1→T2 (D) Not conflict serializable

39. Specialization is _____ process.
(A) top-down (B) bottom up
(C) Both (A) and (B) (D) none of these

40. If D1, D2, .., Dn are domains in a relational model, then the relation is a table, which is a subset of
(A) D1+D2+ … +Dn (B) D1×D2× … ×Dn
(C) D1 ∪ D2 ∪ ⋯ ∪ Dn (D) D1–D2– … –Dn

41.For each value of the ___, the distribution of the dependent variable must be normal.

(A) Independent variable (B)Depended variable

(C)Intermediate variable (D)None of the mentioned above

42. Data Analytics uses ___ to get insights from data.

(A)Statistical figures (B)Numerical aspects

(C)Statistical methods (D)None of the mentioned above

43. Linear Regression is the supervised machine learning model in which the model finds the best fit ___ between the independent and dependent variable.

(A)Linear line (B)Nonlinear line

(C) Curved line (D) All of the mentioned above

44. Amongst which of the following is / are the types of Linear Regression,

(A) Simple Linear Regression (B)Multiple Linear Regression

(C)Both A and B (D)None of the mentioned above

45. Amongst which of the following is / are the true about regression analysis?

(A)Describes associations within the data

(B)Modeling relationships within the data

(C)Answering yes/no questions about the data

(D)All of the mentioned above

46. The process of quantifying data is referred to as ___.

(A)Decoding (B)Structure

(C)Enumeration (D)Coding

47. Data Analysis is a process of,

(A)Inspecting data (B)Data Cleaning

(C) Transforming of data (D)All of the mentioned above

48. Least Square Method uses ___.

(A)Linear polynomial (B)Linear regression

(C) Linear sequence (D)None of the mentioned above

49. What is a hypothesis?

(A)A statement that the researcher wants to test through the data collected in a study

(B)A research question the results will answer

(C)A theory that underpins the study

(D)A statistical method for calculating the extent to which the results could have happened by chance

50. ___ are used when we want to visually examine the relationship between two quantitative variables.

(A)Bar graph (B)Scatterplot

(C)Line graph (D)Pie chart

QNo	Ans. Key	QNo	Ans. Key	QNo	Ans. Key	QNo	Ans. Key	QNo	Ans. Key
1	(A)	11	(C)	21	(C)	31	(C)	41	(A)
2	(B)	12	(B)	22	(D)	32	(A)	42	(C)
3	(B)	13	(A)	23	(C)	33	(D)	43	(A)
4	(A)	14	(C)	24	(B)	34	(C)	44	(C)
5	(D)	15	(D)	25	(D)	35	(B)	45	(B)
6	(A)	16	(D)	26	(B)	36	(D)	46	(C)
7	(A)	17	(A)	27	(B)	37	(D)	47	(D)
8	(D)	18	(D)	28	(C)	38	(B)	48	(B)
9	(A)	19	(B)	29	(A)	39	(C)	49	(A)
10	(C)	20	(D)	30	(B)	40	(D)	50	(A)

ADL332	BIG DATA ANALYTICS LAB	CATEGORY	L	T	P	CREDIT	YEAR OF INTRODUCTION
		PCC	0	0	3	3	2019

Preamble: The purpose of the course is to offer the students a hands-on experience on Big Data concepts using open source technologies such as Hadoop, Map Reduce, Hive, Pig and Apache Spark. The hands-on experience with R Programming language helps in statistical analysis and equip the students with data driven solutions for the next-generation data management. As data continues to grow it is known that via big data solutions, organizations generate insights and make well-informed decisions, discover trends, and improve productivity and the learner will be able to work on and solve data processing problems.

Prerequisite: Fundamental knowledge in Java programming, Statistics and Python and Big Data Analytics

Course Outcomes: At the end of the course, the student should be able to :

CO1	Illustrate the setting up of and Installing Hadoop in one of the three operating modes.**(Cognitive knowledge: Understand)**
CO2	Implement the file management tasks in Hadoop and explore the shell commands **(Cognitive knowledge: Apply)**
CO3	Implement different tasks using Hadoop Map Reduce programming model.**(Cognitive knowledge: Apply)**
CO4	Implement Pig Scripting operations and Spark Application functionalities.**(Cognitive knowledge: Apply)**
CO5	Implement data extraction from files and other sources and perform various data manipulation tasks on them using R Program.**(Cognitive knowledge: Apply)**
CO6	Illustrate the knowledge of R gained to data analytics for real life applications. **(Cognitive knowledge: Understand)**

Mapping of course outcomes with program outcomes

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

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

Assessment Pattern:

Bloom's Category	Continuous Assessment Test(Internal Exam) Marks in percentage	End Semester Examination Marks in percentage
Remember	20	20
Understand	20	20
Apply	60	60
Analyse		
Evaluate		
Create		

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	**75**	**75**	**3 hours**

Continuous Internal Evaluation Pattern:

Attendance : 15 marks
Continuous Evaluation in Lab : 30 marks
Continuous Assessment Test : 15 marks
Viva Voce : 15 marks

Internal Examination Pattern: The marks will be distributed as Algorithm30 marks, Program 20 marks, Output 20 marks and Viva 30 marks. Total 100 marks which will be converted out of 15 while calculating Internal Evaluation marks.

End Semester Examination Pattern: The percentage of marks will be distributed as Algorithm 30 marks, Program 20 marks, Output 20 marks and Viva30 marks. Total 75 marks.

Operating System to Use in Lab : Linux
Compiler/Software to Use in Lab :
Programming Language to Use in Lab : Java, R, Python

Fair Lab Record:
All Students attending the Big Data Lab should have a Fair Record. The fair record should be produced in the University Lab Examination. Every experiment conducted in the lab should be noted in the fair record. For every experiment in the fair record, the right-hand page should contain Experiment Heading, Experiment Number, Date of experiment, Aim of the Experiment and the operations performed on them, Details of experiment including algorithm and result of Experiment. The left-hand page should contain a print out of the code used for experiment and sample output obtained for a set of input.

SYLLABUS

BIG DATA ANALYTICS LAB

* Mandatory

1. Perform setting up and Installing Hadoop in any of the three operating modes: Standalone, Pseudo distributed, Fully distributed.*
2. Explore the various shell commands in Hadoop.
3. Implement the following file management tasks in Hadoop:
   - Adding Files and Directories
   - Retrieving Files
   - Deleting Files
4. Implement a word count program using Map Reduce.
5. Write a R program to find the factorial and check for palindromes.*
6. Write a R program to solve linear regression and make predictions.*
7. Write a R program to solve logistic regression.*
8. Implement statistical operations using R.*
9. Implement a program to find variance, covariance and correlation between different types of attributes.*
10. Implement SVM/Decision tree Classifier.*
11. Implement clustering algorithm.*

12. To explore Hive with its basic commands
13. Write Pig Latin scripts to sort, group, join, project, and filter your data.
14. Install, Deploy and configure Apache Spark.

BIG DATA PROCESSING LAB - PRACTICE QUESTIONS

1. Write a MapReduce Program to retrieve data from documents.
2. Write word count program that only count the words starting with 'a'
3. Write a word count program that only counts the words whose length is longer than 10.
4. Using the structure of the Word Count program, write a Hadoop program that calculates the average word length of all words that start with each character.
5. Implement matrix multiplication with Hadoop Map Reduce
6. Write a Map Reduce program for removing stop words from the given text files.
7. Write a MapReduce Program to count the number of lines in a document.
8. Write Pig Latin script to count the number of occurrences of each word in an input text file.
9. Write a program to simulate Singular Value Decomposition
10. Write a program to simulate PCA.
11. Write a single Spark application that:
    a. Transposes the original Amazon food dataset, obtaining a Pair RDD of the type: user-id – list of the product-ids reviewed by user-id
    b. Counts the frequencies of all the pairs of products reviewed together;
    c. Writes on the output folder all the pairs of products that appear more than once and their frequencies.
    d. The pairs of products must be sorted by frequency..
12. Write a program to implement a stop word elimination problem. Input: A large textual file containing one sentence per line. A small file containing a set of Stop Words (One Stop Word per line) Output: A textual file containing the same sentences of the large input file without the words appearing in the small file
13. Implement matrix multiplication with Map Reduce.
14. Implement basic Pig Latin Scripts based on different scenarios.
15. Implement Frequent Item set algorithm

16. Implement Clustering algorithm
17. Implement Page Rank algorithm
18. Implement Bloom Filter
19. Write a R program to create a sequence of numbers from 20 to 50 and find the mean of numbers from 20 to 60 and sum of numbers from 51 to 91.
20. Write a R program to create a vector which contains 10 random integer values between -50 and +50.
21. Write a R program to find the maximum and the minimum value of a given vector.
22. Write a R program to get the unique elements of a given string and unique numbers of vectors.
23. Write a R program to create a list of random numbers in normal distribution and count occurrences of each value.
24. Write a R program to read the .csv file and display the content.
25. Write a R program to create an array, passing in a vector of values and a vector of dimensions. Also provide names for each dimension.
26. Write a R program to create a simple bar plot of five subjects' marks.
27. Write a R program to compute the sum, mean and product of a given vector element.
28. Write a R program to create a Data Frames which contain details of 5 employees and display the details.

ADD334	MINI PROJECT	CATEGORY	L	T	P	CREDITS
		PWS	0	0	3	2

Preamble: The objective of this course is to apply the fundamental concepts of Artificial Intelligence / Data Science principles for the effective development of an application/research project. Mini project enables the students to boost their skills, widen the horizon of thinking and their ability to resolve real life problems. The students are expected to design and develop a software/hardware project to innovatively solve a real-world problem.

Prerequisite :A sound knowledge in any programming language and Subjects studied up to sixth semester.

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

CO#	CO
CO1	Identify technically and economically feasible problems of social relevance **(Cognitive Knowledge Level: Apply)**
CO2	Identify and survey the relevant literature for getting exposed to related solutions **(Cognitive Knowledge Level: Apply)**
CO3	Perform requirement analysis and identify design methodologies and develop adaptable and reusable solutions of minimal complexity by using modern tools and advanced programming techniques (**Cognitive Knowledge Level: Apply)**
CO4	Prepare technical report and deliver presentation**(Cognitive Knowledge Level: Apply)**
CO5	Apply engineering and management principles to achieve the goal of the project **(Cognitive Knowledge Level: Apply)**

Mapping of course outcomes with program outcomes

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

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

Assessment Pattern

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	75	75	3

Split-up of Continuous Internal Evaluation :

Attendance **10 marks**
Project Guide **15 marks**
Project Report **10 marks**

Evaluation by the Committee (will be evaluating the level of completion and demonstration of functionality/specifications, presentation,

oral examination, work knowledge and involvement) **40 marks**

Split-up of End Semester Examination: The marks will be distributed as

Presentation	: 30 marks
Demonstration	: 20 marks
Viva	: 25 marks.
Total	: 75 marks.

Course Plan

Student Groups with 3 or 4 members should identify a topic of interest in consultation with Faculty/Advisor. Review the literature and gather information pertaining to the chosen topic. State the objectives and develop a methodology to achieve the objectives. Carryout the design/fabrication or develop codes/programs to achieve the objectives. Innovative design concepts, performance, scalability, reliability considerations, aesthetics/ergonomic, user experience and security aspects taken care of in the project shall be given due weight.

The progress of the mini project is evaluated based on a minimum of two reviews. The review committee may be constituted with the Head of the Department or a senior faculty, Mini Project coordinator and project guide as the members. Innovative design concepts, reliability considerations, aesthetics/ergonomic aspects taken care of in the project shall be given due weight. The internal evaluation shall be made based on the progress/outcome of the project, reports and a viva-voce examination, conducted internally by a 3-member committee. A project report is required at the end of the semester. The product/application has to be demonstrated for its full design specifications.

Guidelines for the Report preparation

A bonafide report on mini project shall be submitted within one week after the final presentation. Minimum number of pages should be 40.

- Use Times New Roman font for the entire Report – Chapter / Section Title –Times New Roman 18, Bold; Heading 2 – Times New Roman 16, Bold; Heading 3 – Times New Roman 14,Bold; Body- Times New Roman12, Normal.
- Line Spacing – Between Heading 2 – 3 lines, between lines in paragraph 1.5 lines.
- Alignments – Chapter / Section Title – Center, Heading 2 & 3 should be LeftAligned. Ensure that all body text is paragraph justified.
- Figures & Tables – Ensure that all Figures and Tables are suitably numbered and given proper names/headings. Write figure title under the figure and table title above the table
- Suggestive order of documentation:
  - i. Top Cover
  - ii. Title page
  - iii. Certification page
  - iv. Acknowledgement

v. Abstract
vi. Table of Contents
vii. List of Figures and Tables
viii. Chapters
ix. Appendices, if any
x. References/Bibliography