JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
M. Tech. in DATA SCIENCE
EFFECTIVE FROM ACADEMIC YEAR 2025 - 26 ADMITTED BATCH
COURSE STRUCTURE AND SYLLABUS (R25)

I YEAR I – SEMESTER

Course Code	Course Title	L	T	P	Credits
Professional Core – I	Advanced Data Structures using Python	3	0	0	3
Professional Core – II	Statistical Foundations for Data Science	3	0	0	3
Professional Elective – I	1. Image and Video Processing 2. Advanced Databases 3. Data Wrangling and Visualization	3	0	0	3
Professional Elective- II	1. Data Stream Processing 2. Social Media Analytics 3. Web & Database Security	3	0	0	3
Lab – I	Advanced Data Structures Lab	0	0	4	2
Lab – II	PE-I Lab	0	0	4	2
	Research Methodology & IPR	2	0	0	2
Audit – I	Audit Course – I	2	0	0	0
	Total	**16**	**0**	**8**	**18**

I YEAR II –SEMESTER

Course Code	Course Title	L	T	P	Credits
Professional Core – III	Big Data Technologies	3	0	0	3
Professional Core – IV	Deep Learning	3	0	0	3
Professional Elective – III	1. Natural Language Processing 2. Edge Analytics 3. Enterprise Cloud Concepts	3	0	0	3
Professional Elective – IV	1. Predictive Analytics 2. Mining Massive Datasets 3. Nature Inspired Computing	3	0	0	3
Lab – III	Big Data Analytics Lab	0	0	4	2
Lab – IV	PE-III Lab	0	0	4	2
	Mini Project with Seminar	0	0	4	2
Audit – II	Audit Course – II	2	0	0	0
	Total	14	0	12	18

II YEAR I – SEMESTER

Course Code	Course Title	L	T	P	Credits
Professional Elective – V	1. Digital Forensics 2. Conversational AI 3. Video Analytics	3	0	0	3
Open Elective	Open Elective	3	0	0	3
Dissertation	Dissertation Work Review – I	0	0	18	6
	Total	**6**	**0**	**18**	**12**

II YEAR II - SEMESTER

Course Code	Course Title	L	T	P	Credits
Dissertation	Dissertation Work Review – II	0	0	18	6
Dissertation	Dissertation Viva-Voce	0	0	42	14
	Total	0	0	60	20

Note: For Dissertation Work Review - I, Please refer R25 Academic Regulations.

Audit Course I & II:

1. English for Research Paper Writing
2. Disaster Management
3. Sanskrit for Technical Knowledge
4. Value Education
5. Constitution of India
6. Pedagogy Studies
7. Stress Management by yoga
8. Personality Development Through Life Enlightenment Skills

Open Electives:

1. Statistical Foundations for Data Science
2. Big Data Technologies
3. Predictive Analytics
4. Generative AI

ADVANCED DATA STRUCTURES USING PYTHON (PC-I)

M.Tech DS I Year I Sem.

L	T	P	C
3	0	0	3

Pre-Requisites: UG level course in Data Structures

Course Objectives:

1. Implement Object Oriented Programming concepts in Python.
2. Understand Lists, Dictionaries and Regular expressions in Python.
3. Understanding how searching and sorting is performed in Python.
4. Understanding how linear and non-linear data structures works.
5. To learn the fundamentals of writing Python scripts.

Course Outcomes:

1. Examine Python syntax and semantics and apply Python flow control and functions.
2. Create, run and manipulate Python Programs using core data structures like Lists,
3. Apply Dictionaries and use Regular Expressions.
4. Interpret the concepts of Object-Oriented Programming as used in Python.
5. Master object-oriented programming to create an entire python project using objects and classes

UNIT – I

Oops Concepts- class, object, constructors, types of variables, types of methods. **Inheritance:** single, multiple, multi-level, hierarchical, hybrid, **Polymorphism:** with functions and objects, with class methods, with inheritance, **Abstraction:** abstract classes.

UNIT – II

Data Structures – Definition, Linear Data Structures, Non-Linear Data Structures

Python Specific Data Structures: List, Tuples, Set, Dictionaries, Comprehensions and its Types, Strings, slicing.

UNIT -III

Arrays - Overview, Types of Arrays, Operations on Arrays, Arrays vs List.

Searching -Linear Search and Binary Search.

Sorting - Bubble Sort, Selection Sort, Insertion Sort, Merge Sort, Quick Sort.

UNIT -IV

Linked Lists – Implementation of Singly Linked Lists, Doubly Linked Lists, Circular Linked Lists.

Stacks - Overview of Stack, Implementation of Stack (List & Linked list), Applications of Stack

Queues- Overview of Queue, Implementation of Queue(List & Linked list), Applications of Queues, Priority Queues.

UNIT -V

Graphs -Introduction, Directed vs Undirected Graphs, Weighted vs Unweighted Graphs, Representations, Breadth First Search, Depth First Search.

Trees - Overview of Trees, Tree Terminology, Binary Trees: Introduction, Implementation, Applications. Tree Traversals, Binary Search Trees: Introduction, Implementation, AVL Trees: Introduction, Rotations,
Implementation.

TEXT BOOKS:

1. Data structures and algorithms in python by Michael T. Goodrich
2. Data Structures and Algorithmic Thinking with Python by Narasimha Karumanchi

REFERENCE BOOKS:

1. Hands-On Data Structures and Algorithms with Python: Write complex and powerful code using the latest features of Python 3.7, 2nd Edition by Dr. Basant Agarwal, Benjamin Baka.
2. Data Structures and Algorithms with Python by Kent D. Lee and Steve Hubbard.
3. Problem Solving with Algorithms and Data Structures Using Python by Bradley N Miller and David L.Ranum.
4. Core Python Programming -Second Edition,R. Nageswara Rao, Dreamtech Press

STATISTICAL FOUNDATIONS FOR DATA SCIENCE (PC-II)

M.Tech DS I Year I Sem.

L	T	P	C
3	0	0	3

Prerequisites: Mathematics courses of first year of study.

Course Objectives:

1. The Number Theory basic concepts useful for cryptography etc
2. The theory of Probability, and probability distributions of single and multiple random variables
3. The sampling theory and testing of hypothesis and making inferences
4. Stochastic process and Markov chains.

Course Outcomes: After learning the contents of this course, the student must be able to

1. Apply the number theory concepts to cryptography domain
2. Apply the concepts of probability and distributions to some case studies
3. Correlate the material of one unit to the material in other units
4. Resolve the potential misconceptions and hazards in each topic of study.

UNIT - I

Greatest Common Divisors and Prime Factorization: Greatest common divisors, The Euclidean algorithm, The fundamental theorem of arithmetic, Factorization of integers and the Fermat numbers Congruences: Introduction to congruences, Linear congruences, The Chinese remainder theorem, Systems of linear congruences

UNIT - II

Simple Linear Regression and Correlation: Introduction to Linear Regression, The Simple Linear Regression Model, Least Squares and the Fitted Model, Properties of the Least Squares Estimators, Inferences Concerning the Regression Coefficients, Prediction, Simple Linear Regression Case Study Random Variables and Probability Distributions: Concept of a Random Variable, Discrete Probability Distributions, Continuous Probability Distributions, Statistical Independence. Discrete Probability Distributions: Binomial Distribution, Poisson distribution.

UNIT - III

Continuous Probability Distributions: Normal Distribution, Areas under the Normal Curve, Applications of the Normal Distribution, Normal Approximation to the Binomial, Fundamental Sampling Distributions: Random Sampling, Sampling Distributions, Sampling, Distribution of Means and the Central Limit Theorem, Sampling Distribution of S2, t–Distribution, F Distribution.

UNIT - IV

Estimation & Tests of Hypotheses: Introduction, Statistical Inference, Classical Methods of Estimation. Estimating the Mean, Standard Error of a Point Estimate, Prediction Intervals, Tolerance Limits, Estimating the Variance, Estimating a Proportion for single mean, Difference between Two Means, between Two Proportions for Two Samples and Maximum Likelihood Estimation.

UNIT - V

Stochastic Processes and Markov Chains: Introduction to Stochastic processes- Markov process. Transition Probability, Transition Probability Matrix, First order and Higher order Markov process, n step transition probabilities, Markov chain, Steady state condition, Markov analysis.

TEXT BOOKS:

1. Kenneth H. Rosen, Elementary number theory & its applications, sixth edition, Addison Wesley, ISBN 978 0-321-50031-1
2. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, Probability & Statistics for Engineers & Scientists, 9th Ed. Pearson Publishers.
3. S. D. Sharma, Operations Research, Kedarnath and Ramnath Publishers, Meerut, Delhi

REFERENCE BOOKS:

1. S C Gupta and V K Kapoor, Fundamentals of Mathematical statistics, Khanna publications
2. T.T. Soong, Fundamentals of Probability And Statistics For Engineers, John Wiley & Sons Ltd, 2004.
3. Sheldon M Ross, Probability and statistics for Engineers and scientists, Academic Press.

IMAGE AND VIDEO PROCESSING (PROFESSIONAL ELECTIVE -I)

M.Tech DS I Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives:

1. Comprehend the image processing fundamentals and enhancement techniques in spatial and frequency domain.
2. Describe the color image fundamentals, models and various restoration techniques.
3. Design and Analyze the image compression systems.
4. Outline the various image segmentation and morphology operations.
5. Comprehend the basics of video processing and video coding.

Course Outcomes:

1. Understand theory and models in Image and Video Processing.
2. Explain the need of spatial and frequency domain techniques for image compression.
3. Comprehend different methods, models for video processing and motion estimation.
4. Illustrate quantitative models of image and video segmentation.
5. Apply the process of image enhancement for optimal use of resources.

Unit I

Digital image fundamentals: A simple image formation model, Image sampling and quantization, Some basic relationships between pixels, Basic intensity transformation functions, Sampling and fourier transform of sampled functions, The discrete fourier transform of one variable, Extensions to functions of two variables (2-D discrete fourier transform, Properties of 2-D DFT and IDFT, 2-D Discrete Convolution Theorem

Unit II

Image Enhancement (spatial domain): Histogram processing, Fundamentals of spatial filtering, Smoothing spatial filters, Sharpening spatial filters, The Laplacian-use of second order derivative for image sharpening, The Gradient-use of first order derivative for image sharpening
Image Enhancement (frequency domain): Basics of filtering in frequency domain, Image smoothing using lowpass frequency domain filters, Image sharpening using highpass filters

Unit III

Image restoration: Noise Models, Restoration in the presence of noise only – Spatial filters, Periodic noise reduction using Frequency domain filtering, Estimating the degradation function, inverse filtering, Minimum Least square error filtering, constrained least square filters
Wavelet and Multi resolution processing: Matrix-based transform, Walsh-Hadamard Transform, Slant transform, Haar transform

Unit IV

Image compression: Lossy and lossless compression schemes: Huffman coding, Run-length coding, Arithmetic coding, Block transform coding, JPEG
Image Morphology: Fundamental operations, Morphological Algorithms
Image segmentation: Point, Line and Edge detection, Canny edge detection, Hough Transform, Edge linking, Thresholding, Region-based segmentation, Pixel-based segmentation.

Unit V

Feature Extraction: Boundary preprocessing, Boundary feature descriptor, Region feature descriptor, Principal components as feature descriptor, Whole image feature
Video Processing: Video Formats, Video Enhancement and Restoration, Video Segmentation

TEXT BOOKS:

1. Digital Image Processing, R. C. Gonzalez and R. E. woods, Pearson Education.
2. Handbook of Image and Video Processing, AL Bovik, Academic Press.

REFERENCES:

1. Digital Image Processing and Analysis, B. Chanda and D. Dutta Mazumdar, PHI.
2. Digital Image Processing, W. K. Pratt, Wiley-Interscience.
3. Fundamentals of Digital Image Processing, A. K. Jain, Pearson India Education.
4. Pattern Classification and Scene Analysis, R. O. Duda and P. E. Hart, Wiley.

ADVANCED DATABASES (PROFESSIONAL ELECTIVE -I)

M.Tech DS I Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives:
Knowledge on concepts of Distributed Databases, Object-Based Databases, advanced database models

Course Outcomes:
1. Understand Database system Architectures and parallel databases
2. Analyze transactions, Concurrency Control in Distributed Databases
3. Understand the importance of Data Warehousing and Mining
4. Illustrate concepts of object based databases

Unit I Database System Architectures
Centralized and Client –Server Architectures, Server System Architectures, Parallel Systems, Distributed Systems, Network Types
Parallel Databases: Introduction, I/O Parallelism, Interquery Parallelism, Intra Query Parallelism, Intraoperation Parallelism, Interoperation Parallelism, Query Optimization, Design of Parallel Systems, Parallelism on Multicore Processors

Unit II
Distributed Databases
Homogeneous and Heterogeneous Databases, Distributed Data Storage, Distributed Transactions, Commit Protocols, Concurrency Control in Distributed Databases, Availability, Distributed Query Processing, Heterogeneous Distributed Databases, Cloud-Based Databases, Directory Systems

Unit III Data Warehousing and Mining
Decision-Support Systems, Data Warehousing, Data Mining, Classification, Association Rules, Other Types of Associations, Clustering, Other Forms of Data Mining

Unit IV Object-Based Databases
Introduction, Complex Data Types, Structured Types and Inheritance in SQL, Table Inheritance, Array and Multiset Types in SQL, Object-Identity and Reference Types in SQL, Implementing O-R Features, Persistent Programming Languages, Object-Relational Mapping, Object-Oriented versus Object-Relational

Unit V XML
Motivation, Structure of XML Data, XML Document Schema, Querying and Transformation, Application Program Interfaces to XML, Storage of XML Data, XML Applications
Advanced database models and applications: Active Database Concepts and Triggers, Temporal database concepts, Spatial database concepts, Multimedia database concepts, Deductive databases

TEXT BOOKS:
1. Abraham Silberschatz, Henry F. Korth, S. Sudarshan Database System Concepts, Sixth Edition
2. Ramez Elmasri, Shamkant B. Navathe, Database systems- Models, Languages, Design and Application Programming

DATA WRANGLING AND DATA VISUALIZATION (PROFESSIONAL ELECTIVE -I)

M.Tech DS I Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives:

1. To learn data wrangling techniques
2. To introduce visual perception and core skills for visual analysis

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

1. Perform data wrangling
2. Explain principles of visual perception
3. Apply core skills for visual analysis
4. Apply visualization techniques for various data analysis tasks
5. Evaluate visualization techniques

Unit I :
Data Wrangling: Need of data cleanup, data clean up basics – formatting, outliers, duplicates, Normalizing and standardizing data.

Unit II :
Introduction of visual perception, visual representation of data, Gestalt principles, information overloads.
Creating visual representations, visualization reference model, visual mapping, visual analytics, Design of visualization applications.

Unit III : Classification of visualization systems, Interaction and visualization techniques misleading, Visualization of one, two and multi-dimensional data, text and text documents.

Unit IV : Visualization of groups, trees, graphs, clusters, networks, software, Metaphorical visualization

Unit V : Visualization of volumetric data, vector fields, processes and simulations, Visualization of maps, geographic information, GIS systems, collaborative visualizations, evaluating visualizations

TEXT BOOKS:

1. Jacqueline Kazil and Katharine Jarmul, Data Wrangling with Python: Tips and Tools to Make Your Life Easier, O'Reilly.
2. Ward, Grinstein Keim, Interactive Data Visualization: Foundations, Techniques, and Applications. Natick A K Peters, Ltd.

REFERENCES:

1. E. Tufte, The Visual Display of Quantitative Information, Graphics Press.

DATA STREAM PROCESSING (PROFESSIONAL ELECTIVE - III)

M.Tech DS I Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives

1. Introduce the student to use cases of stream processing, the data stream model
2. Present algorithmic techniques for stream processing, including map-and-reduce.
3. Present current techniques on monitoring parallel and distributed streams.
4. Implement streams in the continuous distributed monitoring model, and the parallel streaming model.
5. Provide practical perspective on building software for stream processing

Course Outcomes

1. Explain the need for stream processing
2. Comprehend the architectures of stream processing.
3. Explain and run Distributed Processing and Resilience Model
4. Design effective streaming solutions using Structured Streaming
5. Design effective streaming solutions using Spark Streaming

UNIT-I

Introduction To Stream Processing Model

Fundamentals of Stream Processing: What Is Stream Processing? Examples of Stream Processing- Scaling Up Data Processing- Distributed Stream Processing- Introducing Apache Spark. Stream-Processing Model: Sources and Sinks- Immutable Streams Defined from One Another Transformations and Aggregations- Window Aggregations - Stateless and Stateful Processing- The Effect of Time.

UNIT-II

Streaming Architectures

Components of a Data Platform- Architectural Models- The Use of a Batch-Processing Component in a Streaming Application- Referential Streaming Architectures- Streaming Versus Batch Algorithms. Apache Spark as a Stream-Processing Engine: Spark's Memory Usage- Understanding Latency- Throughput Oriented Processing- Fast Implementation of Data Analysis

UNIT-III

Distributed Processing And Resilience Model

Spark's Distributed Processing Model: Running Apache Spark with a Cluster Manager- Spark's Own Cluster Manager - Resilience and Fault Tolerance in a Distributed System- Data Delivery Semantics- Micro Batching and One-Element-at-a-Time - Bringing Micro batch and One-Record-at a- Time Closer Together- Dynamic Batch Interval- Structured Streaming Processing Model. Spark's Resilience Model: Resilient Distributed Datasets in Spark - Spark Components - Spark's Fault-Tolerance Guarantees.

UNIT-IV

Structured Streaming

Introducing Structured Streaming- The Structured Streaming Programming Model – Structured Streaming in Action – Structured Streaming Sources – Structured Streaming Sinks - Event Time–Based Stream Processing.

UNIT-V

Spark Streaming

Introducing Spark Streaming - The Spark Streaming Programming Model - The Spark Streaming Execution Model - Spark Streaming Sources - Spark Streaming Sinks - Time-Based Stream Processing- Working with Spark SQL – Checkpointing - Monitoring Spark Streaming- Performance Tuning.

TEXT BOOKS:

1. Gerard Maas and François Garillot , "Stream Processing with Apache Spark: Mastering Structured Streaming and Spark Streaming", O'Reilly, 2019.

REFERENCE BOOKS:

1. Henrique C. M. Andrade, Buğra Gedik and Deepak S. Turaga, “Fundamentals of Stream Processing: Application Design, Systems, and Analytics”, Cambridge University Press, 2014.
2. Bryon Ellis, “Real-Time Analytics: Techniques to Analyze and Visualize Streaming Data”, Wiley, 1st edition, 2014.
3. Anindita Basak, Krishna Venkataraman, Ryan Murphy, Manpreet Singh, “Stream Analytics with Microsoft Azure”, Packt Publishing, December 2017.

Web references:

1. https://github.com/stream-processing-with-spark

E -Text Books:

1. https://www.edx.org/course/processing-real-time-data-streams-in-azure
2. https://www.coursera.org/learn/big-data-integration-processing

SOCIAL MEDIA ANALYTICS (PROFESSIONAL ELECTIVE - II)

M.Tech DS I Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives: Knowledge on social media and its analytics

Course Outcomes

1. Understanding characteristics and types of social media
2. Knowledge on layers of social media analytics
3. Apply text analysis tools on social media data
4. Understand the significance of action analytics
5. Detect viral topics on social media(YouTube)

Unit I

Introduction To Social Media

World Wide Web, Web 1.0, Web 2.0, Web 3.0, Social Media, Core Characteristics Of Social Media, Types Of Social Media, Social Networking Sites, Using Facebook For Business Purposes, Content Communities

Unit- II

Social Media Analytics Overview

Purpose Of Social Media Analytics, Social Media Vs. Traditional Business Analytics, Seven Layers Of Social Media Analytics, Types Of Social Media Analytics, Social Media Analytics Cycle, Challenges To Social Media Analytics, Social Media Analytics Tools.

Case Study: The Underground Campaign That Scored Big

Unit III

Social Media Text Analytics

Types Of Social Media Text, Purpose Of Text Analytics, Steps In Text Analytics, Social Media Text Analysis Tools.

Case Study: Tapping Into Online Customer Opinions

Unit IV

Social Media Actions Analytics

Introduction To Actions Analytics, Common Social Media Actions, Actions Analytics Tools.

Case Study: Cover-More Group

Unit V

Social Media Hyperlink Analytics

Types Of Hyperlinks, Hyperlink Analytics, Types Of Hyperlink Analytics, Hyperlink Analytics Tools.

Case Study: Hyperlinks And Viral YouTube Videos

TEXT BOOKS

1. Seven Layers Of Social Media Analytics Mining Business Insights From Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine, And Location Data By Gohar F. Khan Isbn: 1507823207, Isbn-13: 9781507823200

REFERENCE BOOKS:

1. Social Media Analytics: Techniques And Insights For Extracting Business Value Out Of Social Media By Matthew Ganis, Avinash Kohirkar, Pearson Education.
2. Social Media Analytics: Effective Tools for Building, Interpreting, and Using Metrics, Marshall Sponder, MGH.
3. Big Data And Analytics,Seema Acharya, Subhasinin Chellappan, Wiley Publications.
4. Big Data, Black Book^{tm} , Dreamtech Press, 2015 Edition.

WEB & DATABASE SECURITY (PROFESSIONAL ELECTIVE - II)

M.Tech DS I Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives

1. Give an Overview of information security
2. Give an overview of Access control of relational databases

Course Outcomes: Students should be able to

1. Understand the Web architecture and applications
2. Understand client side and server side programming
3. Understand how common mistakes can be bypassed and exploit the application
4. Identify common application vulnerabilities

UNIT - I

The Web Security,The Web Security Problem ,Risk Analysis and Best Practices Cryptography and the Web : Cryptography and Web Security, Working Cryptographic Systems and Protocols , Legal Restrictions on Cryptography ,Digital Identification

UNIT - II

The Web's War on Your Privacy, Privacy-Protecting Techniques, Backups and anti-theft, Web Server Security, Physical Security for Servers, Host Security for Servers, Securing Web Applications

UNIT - III

Database Security: Recent Advances in Access Control, Access Control Models for XML, Database Issues in Trust Management and Trust Negotiation, Security in Data Warehouses and OLAP Systems

UNIT - IV

Security Re-engineering for Databases: Concepts and Techniques, Database Watermarking for Copyright Protection, Trustworthy Records Retention, Damage Quarantine and Recovery in Data Processing Systems, Hippocratic Databases: Current Capabilities and

UNIT - V

Future Trends Privacy in Database Publishing: A Bayesian Perspective, Privacy-enhanced Location-based Access Control, Efficiently Enforcing the Security and Privacy Policies in a Mobile Environment

TEXT BOOKS:

1. Web Security, Privacy and Commerce Simson Garfinkel, Gene Spafford, O' Reilly.
2. Handbook on Database security applications and trends Michael Gertz, Sushil Jajodia

REFERENCES:

1. Andrew Hoffman, Web Application Security: Exploitation and Countermeasures for Modern Web Applications, O'reilly
2. Jonathan LeBlanc Tim Messerschmidt, Identity and Data Security for Web Development - Best Practices, O'reilly
3. McDonald Malcolm, Web Security For Developers, No Starch Press, US

ADVANCED DATA STRUCTURES USING PYTHON LAB (LAB - I)

M.Tech DS I Year I Sem.

L	T	P	C
0	0	4	2

Pre-Requisites: UG level course in Data Structures

COURSE OBJECTIVES: This course will enable students to
1. Implement Object Oriented Programming concepts in Python.
2. Understand Lists, Dictionaries and Regular expressions in Python.
3. Understanding how searching and sorting is performed in Python.
4. Understanding how linear and non-linear data structures works.
5. To learn the fundamentals of writing Python scripts.

COURSE OUTCOMES: The students should be able to:
1. Examine Python syntax and semantics and apply Python flow control and functions.
2. Create, run and manipulate Python Programs using core data structures like Lists,
3. Apply Dictionaries and use Regular Expressions.
4. Interpret the concepts of Object-Oriented Programming as used in Python.
5. Master object-oriented programming to create an entire python project using objects and classes

List of Experiments:
1. Write a Python program for class, Flower, that has three instance variables of type str, int, and float, that respectively represent the name of the flower, its number of petals, and its price. Your class must include a constructor method that initializes each variable to an appropriate value, and your class should include methods for setting the value of each type, and retrieving the value of each type.
2. Develop an inheritance hierarchy based upon a Polygon class that has abstract methods area( ) and perimeter( ). Implement classes Triangle, Quadrilateral, Pentagon, that extend this base class, with the obvious meanings for the area( ) and perimeter( ) methods. Write a simple program that allows users to create polygons of the various types and input their geometric dimensions, and the program then outputs their area and perimeter.
3. Write a python program to implement Method Overloading and Method Overriding.
4. Write a Python program to illustrate the following comprehensions:
   a) List Comprehensions b) Dictionary Comprehensions c) Set Comprehensions d) Generator Comprehensions
5. Write a Python program to generate the combinations of n distinct objects taken from the elements of a given list. Example: Original list: [1, 2, 3, 4, 5, 6, 7, 8, 9] Combinations of 2 distinct objects: [1, 2] [1, 3] [1, 4] [1, 5] .... [7, 8] [7, 9] [8, 9].
6. Write a python program for Linear Search and Binary search
7. Write a python program to implement Bubble Sort and Selection Sort.
8. Write a python program to implement Merge sort and Quicksort.
9. Write a python program to implement Stacks and Queues.
10. Write a python program to implement a Singly Linked List.
11. Write a python program to implement a Doubly Linked list.
12. Write a python program to implement Binary Search Tree.

TEXTBOOKS:
1. Data structures and algorithms in python by Michael T. Goodrich
2. Data Structures and Algorithmic Thinking with Python by Narasimha Karumanchi

REFERENCE BOOKS:
1. Hands-On Data Structures and Algorithms with Python: Write complex and powerful code using the latest features of Python 3.7, 2nd Edition by Dr. Basant Agarwal, Benjamin Baka.
2. Data Structures and Algorithms with Python by Kent D. Lee and Steve Hubbard.
3. Problem Solving with Algorithms and Data Structures Using Python by Bradley N Miller and David L.Ranum.
4. Core Python Programming -Second Edition,R. Nageswara Rao, Dreamtech Press

IMAGE AND VIDEO PROCESSING LAB (LAB - II)

M.Tech DS I Year I Sem.

L	T	P	C
0	0	4	2

Course Objectives:

1. Comprehend the image processing fundamentals and enhancement techniques in spatial and frequency domain.
2. Describe the color image fundamentals, models and various restoration techniques.
3. Design and Analyze the image compression systems.
4. Outline the various image segmentation and morphology operations.
5. Comprehend the basics of video processing and video coding.

Course Outcomes:

1. Understand theory and models in Image and Video Processing.
2. Explain the need of spatial and frequency domain techniques for image compression.
3. Comprehend different methods, models for video processing and motion estimation.
4. Illustrate quantitative models of image and video segmentation.
5. Apply the process of image enhancement for optimal use of resources.

The students will be given 8-10 programming/simulation/projects/assignments based on the syllabus as mentioned below:

List of Experiments:

1. Implement any four Image Enhancement Algorithms
2. Frequency Domain Filtering Algorithms.
3. Image Restoration Algorithms.
4. Color Image Processing Algorithms
5. Image compression techniques
6. Morphological Image Processing Methods
7. Image Segmentation Algorithms
8. Spatial Filtering Algorithms for image enhancement and segmentation
9. Object Recognition algorithms
10. Video restoration techniques
11. Video segmentation techniques
12. Video compression techniques

ADVANCED DATABASES LAB (LAB - II)

M.Tech DS I Year I Sem.

L	T	P	C
0	0	4	2

Course Objectives:
Knowledge on concepts of Distributed Databases, Object-Based Databases, advanced database models

Course Outcomes:
1. Understand Database system Architectures and parallel databases
2. Analyze transactions, Concurrency Control in Distributed Databases
3. Understand the importance of Data Warehousing and Mining
4. Illustrate concepts of object based databases

List of Experiments
1. Write a program to implement RDBMS - Cursors, Triggers
2. Write a Program to implement Range Partitioning sort.
3. Write a program to implement parallel hash join
4. Write a program to implement parallel nested join loop
5. Write a program to implement parallelize duplicate elimination by partitioning the tuples
6. Perform data fragmentation of distributed data(Horizontal, Vertical, Hybrid fragmentation)
7. Implement deadlock detection in distributed databases
8. Implement Semi Join algorithm.
9. DataCube Implementation - Aggregation
10. Perform data Integration - Extraction, Transformation, Loading
11. Implement any one classifier
12. Implement vector space models for Text corpus
13. Demonstrate type inheritance, table inheritance in object based databases
14. Write queries in XQueries on DTD
15. Write queries in SQL/XML to convert University data - XML Schema

TEXT BOOKS:
1. Abraham Silberschatz, Henry F. Korth, S. Sudarshan Database System Concepts, Sixth Edition
2. Ramez Elmasri, Shamkant B. Navathe, Database systems- Models, Languages, Design and Application Programming.

DATA WRANGLING AND DATA VISUALIZATION LAB (LAB - II)

M.Tech DS I Year I Sem.

L	T	P	C
0	0	4	2

Course Objectives:

1. To learn data wrangling techniques
2. To introduce visual perception and core skills for visual analysis

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

1. Perform data wrangling
2. Explain principles of visual perception
3. Apply core skills for visual analysis
4. Apply visualization techniques for various data analysis tasks
5. Evaluate visualization techniques

Implement the following experiments using Python

1. Find missing values and perform data imputation.
2. Find outliers in a chosen dataset.
3. Methods to handle duplicate data.
4. Perform data normalization
5. Explore 2-D charts
   i. Clustered bar charts
   ii. Connected dot plots
   iii. Pictograms
   iv. Bubble charts
   v. Radar charts
   vi. Polar charts
   vii. Range chart
   viii. Box-and-whisker plots
   ix. Univariate scatter plots
   x. histograms word cloud
   xi. Pie chart
   xii. Waffle chart
   xiii. Stacked bar chart
   xiv. Tree map.
6. Multi-dimensional data visualization
7. Graph data visualization

TEXT BOOKS:

1. Jacqueline Kazil and Katharine Jarmul, Data Wrangling with Python: Tips and Tools to Make Your Life Easier, O'Reilly
2. Andy Kirk, Data Visualization A Handbook for Data Driven Design, Sage Publications, 2016
3. Philipp K. Janert, Gnuplot in Action, Understanding Data with Graphs, Manning Publications, 2010.

RESEARCH METHODOLOGY & IPR

M.Tech DS I Year I Sem.

L	T	P	C
2	0	0	2

Prerequisite: None

Course Objectives:

1. To understand the research problem
2. To know the literature studies, plagiarism and ethics
3. To get the knowledge about technical writing
4. To analyze the nature of intellectual property rights and new developments
5. To know the patent rights

Course Outcomes: At the end of this course, students will be able to

1. Understand research problem formulation.
2. Analyze research related information
3. Follow research ethics
4. Understand that today's world is controlled by Computer, Information Technology, but tomorrow world will be ruled by ideas, concept, and creativity.
5. Understanding that when IPR would take such important place in growth of individuals & nation, it is needless to emphasis the need of information about Intellectual Property Right to be promoted among students in general & engineering in particular.
6. Understand that IPR protection provides an incentive to inventors for further research work and investment in R & D, which leads to creation of new and better products, and in turn brings about, economic growth and social benefits.

UNIT-I:

Meaning of research problem, Sources of research problem, Criteria Characteristics of a good research problem, Errors in selecting a research problem, Scope and objectives of research problem. Approaches of investigation of solutions for research problem, data collection, analysis, interpretation, Necessary instrumentations

UNIT-II:

Effective literature studies approaches, analysis, Plagiarism, Research ethics

UNIT-III:

Effective technical writing, how to write report, Paper Developing a Research Proposal, Format of research proposal, a presentation and assessment by a review committee

UNIT-IV:

Nature of Intellectual Property: Patents, Designs, Trade and Copyright. Process of Patenting and Development: technological research, innovation, patenting, development. International Scenario: International cooperation on Intellectual Property. Procedure for grants of patents, Patenting under PCT.

UNIT-V:

Patent Rights: Scope of Patent Rights. Licensing and transfer of technology. Patent information and databases. Geographical Indications. New Developments in IPR: Administration of Patent System. New developments in IPR; IPR of Biological Systems, Computer Software etc. Traditional knowledge Case Studies, IPR and IITs.

TEXT BOOKS:

1. Stuart Melville and Wayne Goddard, "Research methodology: an introduction for science & engineering students"
2. C.R. Kothari, Research Methodology, methods & techniques, 2nd edition, New age International publishers

REFERENCES:

1. Ranjit Kumar, 2nd Edition, "Research Methodology: A Step by Step Guide for beginners"
2. Halbert, "Resisting Intellectual Property", Taylor & Francis Ltd ,2007.
3. Mayall, "Industrial Design", McGraw Hill, 1992.

4. Niebel, “Product Design”, McGraw Hill, 1974.
5. Asimov, “Introduction to Design”, Prentice Hall, 1962.
6. Robert P. Merges, Peter S. Menell, Mark A. Lemley, “Intellectual Property in New Technological Age”, 2016.
7. T. Ramappa, “Intellectual Property Rights Under WTO”, S. Chand, 2008

BIG DATA TECHNOLOGIES (PC - III)

M.Tech DS I Year II Sem.

L	T	P	C
3	0	0	3

Course Objectives

1. The purpose of this course is to provide the students with knowledge of Big data Analytics principles and techniques.
2. This course is also designed to give an exposure of the frontiers of Big data Analytics

Courses Outcomes

1. Ability to explain the foundations, definitions, and challenges of Big Data and various Analytical tools.
2. Ability to program using HADOOP and Map reduce, NOSQL
3. Ability to understand the importance of Big Data in Social Media and Mining.

Unit I

Getting an Overview of Big Data

Big Data, History of Data Management – Evolution of Big Data, Structuring Big Data, Elements of Big Data, Big Data Analytics, Careers in Big Data, Future of Big Data

Technologies for Handling Big Data

Distributed and Parallel Computing for Big Data, Introducing Hadoop, Cloud Computing and Big Data, In-Memory Computing Technology for Big Data.

Unit II

Understanding Hadoop Ecosystem

Hadoop Ecosystem, Hadoop Distributed File System, MapReduce, Hadoop YARN, Hbase, Hive, Pig and Pig Latin, Sqoop, ZooKeeper, Flume, Oozie

Understanding MapReduce Fundamentals and HBase

The MapReduce Framework, Techniques to Optimize MapReduce Jobs, Uses of MapReduce, Role of HBase in Big Data Processing

Unit III

Exploring Hive

Introducing Hive, Getting Started with Hive, Data Types in Hive, Built-In Functions in Hive, Hive DDL, Data Manipulation in Hive, Data Retrieval Queries, Using JOINS in Hive

Analyzing Data with Pig

Introducing Pig, Running Pig, Getting Started with Pig Latin, Working with Operators in Pig, Working with Functions in Pig

Unit IV

Using Oozie

Introducing Oozie, Installing and Configuring Oozie, Understanding the Oozie Workflow, Oozie Coordinator, Oozie Bundle, Oozie Parameterization with EL, Oozie Job Execution Model, Accessing Oozie, Oozie SLA

NoSQL Data Management

Introduction to NoSQL, Aggregate Data Models, Key Value Data Model, Document Databases, Relationships, Graph Databases, Schema-Less Databases, Materialized Views, Distribution Models, Sharding, MapReduce Partitioning and Combining, Composing MapReduce Calculations

Unit V

Zookeeper: Installing and Running ZooKeeper, An Example, Group Membership in ZooKeeper, Creating the Group, Joining a Group, Listing Members in a Group, The ZooKeeper Service, Data Model, Operations, Implementation, Consistency, Sessions, Building Applications with ZooKeeper, A Configuration, Service, The Resilient ZooKeeper Application, A Lock Service, More Distributed Data Structures and Protocols, ZooKeeper in Production

Sqoop: Getting Sqoop, Sqoop Connectors, A Sample Import, Generated Code, Imports: A Deeper Look, Working with Imported Data, Importing Large Objects, Performing an Export, Exports: A Deeper Look

TEXT BOOKS:

1. Big data, blackbook, Dream Tech Press, 2015

2. Hadoop: The Definitive Guide, Tom White, 3rd Edition, O'Reilly Media, 2012.

REFERENCES:

1. Big Data Analytics, Seema Acharya, Subhashini Chellappan, Wiley 2015.
2. Simon Walkowiak, Big Data Analytics with R, Packt Publishing, ISBN: 9781786466457
3. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business, Michael Minelli, Michehe Chambers, 1st Edition, Ambiga Dhiraj, Wiley CIO Series, 2013.
4. Big Data Analytics: Disruptive Technologies for Changing the Game, Arvind Sathi, 1st Edition, IBM Corporation, 2012.

DEEP LEARNING (PC - IV)

M.Tech DS I Year II Sem.

L	T	P	C
3	0	0	3

Course Objectives:

1. To understand complexity of Deep Learning algorithms and their limitations
2. To be capable of performing experiments in Deep Learning using real-world data.

Course Outcomes:

1. Implement deep learning algorithms, understand neural networks and traverse the layers of data
2. Learn topics such as convolutional neural networks, recurrent neural networks, training deep networks and high-level interfaces
3. Understand applications of Deep Learning to Computer Vision
4. Understand and analyze Applications of Deep Learning to NLP

UNIT -I

Introduction: Feed forward Neural networks, Gradient descent and the back-propagation algorithm, Unit saturation, the vanishing gradient problem, and ways to mitigate it. RelU Heuristics for avoiding bad local minima, Heuristics for faster training, Nestors accelerated gradient descent, Regularization, Dropout

UNIT -II

Convolutional Neural Networks: Architectures, convolution/pooling layers, Recurrent Neural Networks: LSTM, GRU, Encoder Decoder architectures. Deep Unsupervised Learning: Auto encoders, Variational Auto-encoders, Adversarial Generative Networks, Auto-encoder and DBM Attention and memory models, Dynamic Memory Models

UNIT-III

Applications of Deep Learning to Computer Vision: Image segmentation, object detection, automatic image captioning, Image generation with Generative adversarial networks, video to text with LSTM models, Attention Models for computer vision tasks

UNIT -IV

Applications of Deep Learning to NLP: Introduction to NLP and Vector Space Model of Semantics, Word Vector Representations: Continuous Skip-Gram Model, Continuous Bag-of-Words model (CBOW), Glove, Evaluations and Applications in word similarity

UNIT -V

Analogy reasoning: Named Entity Recognition, Opinion Mining using Recurrent Neural Networks: Parsing and Sentiment Analysis using Recursive Neural Networks: Sentence Classification using Convolutional Neural Networks, Dialogue Generation with LSTMs

TEXT BOOKS:

1. Deep Learning by Ian Good fellow, Yoshua Bengio and Aaron Courville, MIT Press.
2. The Elements of Statistical Learning by T. Hastie, R. Tibshirani, and J. Friedman, Springer.
3. Probabilistic Graphical Models. Koller, and N. Friedman, MIT Press.

REFERENCES:

1. Bishop, C., M., Pattern Recognition and Machine Learning, Springer, 2006.
2. Yegnanarayana, B., Artificial Neural Networks PHI Learning Pvt. Ltd, 2009.
3. Golub, G., H., and Van Loan, C., F., Matrix Computations, JHU Press,2013.
4. Satish Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill Education, 2004.

NATURAL LANGUAGE PROCESSING (PROFESSIONAL ELECTIVE - III)

M.Tech DS I Year II Sem.

L	T	P	C
3	0	0	3

Prerequisites:

1. Data structures, finite automata and probability theory.

Course Objectives:

1. Introduction to some of the problems and solutions of NLP and their relation to linguistics and statistics.

Course Outcomes:

1. Show sensitivity to linguistic phenomena and an ability to model them with formal grammars.
2. Understand and carry out proper experimental methodology for training and evaluating empirical NLP systems
3. Able to manipulate probabilities, construct statistical models over strings and trees, and estimate parameters using supervised and unsupervised training methods.
4. Able to design, implement, and analyze NLP algorithms Able to design different language modeling Techniques.
5. Able to design different language modeling Techniques.

UNIT - I

Finding the Structure of Words: Words and Their Components, Issues and Challenges, Morphological Models

Finding the Structure of Documents: Introduction, Methods, Complexity of the Approaches, Performances of the Approaches

UNIT - II

Syntax Analysis: Parsing Natural Language, Treebanks: A Data-Driven Approach to Syntax, Representation of Syntactic Structure, Parsing Algorithms, Models for Ambiguity Resolution in Parsing, Multilingual Issues

UNIT - III

Semantic Parsing: Introduction, Semantic Interpretation, System Paradigms, Word Sense Systems, Software.

UNIT - IV

Predicate-Argument Structure, Meaning Representation Systems, Software.

UNIT - V

Discourse Processing: Cohesion, Reference Resolution, Discourse Cohesion and Structure Language Modeling: Introduction, N-Gram Models, Language Model Evaluation, Parameter Estimation, Language Model Adaptation, Types of Language Models, Language-Specific Modeling Problems, Multilingual and Cross Lingual Language Modeling

TEXT BOOKS:

1. Multilingual natural Language Processing Applications: From Theory to Practice – Daniel M. Bikel and Imed Zitouni, Pearson Publication
2. Natural Language Processing and Information Retrieval: Tanvier Siddiqui, U.S. Tiwary

REFERENCES:

1. Speech and Natural Language Processing - Daniel Jurafsky & James H Martin, Pearson Publications

EDGE ANALYTICS (Professional Elective - III)

M.Tech DS I Year II Sem.

L	T	P	C
3	0	0	3

Prerequisites

A basic knowledge of "Python Programming"

Course Objectives

1. The aim of the course is to introduce the fundamentals of Edge Analytics
2. The course gives an overview of – Architectures, Components, Communication Protocols and tools used for Edge Analytics

Course Outcomes

1. Understand the concepts of Edge Analytics, both in theory and in practical application
2. Demonstrate a comprehensive understanding of different tools used at edge analytics
3. Formulate, Design and Implement the solutions for real world edge analytics

UNIT-I

Introduction to Edge Analytics

What is edge analytics, Applying and comparing architectures, Key benefits of edge analytics, Edge analytics architectures, Using edge analytics in the real world.

UNIT-II

Basic edge analytics components, Connecting a sensor to the ESP-12F microcontroller, KOM-MICS smart factory platform, Communications protocols used in edge analytics, Wi-Fi communication for edge analytics, Bluetooth for edge analytics communication, Cellular technologies for edge analytics communication, Long-distance communication using LoRa and Signfox for edge analytics.

UNIT-III

Working with Microsoft Azure IoT Hub, Cloud Service providers, Microsoft Azure, Exploring the Azure portal, Azure ioT Hub, Using the Raspberry Pi with Azure IoT edge, Connecting our Raspberry Pi edge device, adding a simulated temperature sensor to our edge device.

UNIT-IV

Using Micropython for Edge Analytics, Understanding Micropython, Exploring the hardware that runs MicroPython, Using MicroPython for an edge analytics application, Using edge intelligence with microcontrollers, Azure Machine Learning designer, Azure IoT edge custom vision.

UNIT-V

Designing a Smart Doorbell with Visual Recognition setting up the environment, Writing the edge code, creating the Node-RED dashboard, Types of attacks against our edge analytics applications, Protecting our edge analytics applications

TEXT BOOK:

1. Hands-On Edge Analytics with Azure IoT: Design and develop IoT applications with edge analytical solutions including Azure IoT Edge by Colin Dow

REFERENCE BOOKS:

1. Learn Edge Analytics - Fundamentals of Edge Analytics: Automated analytics at source using Microsoft Azure by Ashish Mahajan

ENTERPRISE CLOUD CONCEPTS (PROFESSIONAL ELECTIVE - III)

M.Tech DS I Year II Sem.

L	T	P	C
3	0	0	3

Course Objectives:
Knowledge on significance of cloud computing and its fundamental concepts and models.

Course Outcomes:
1. Understand importance of cloud architecture
2. Illustrating the fundamental concepts of cloud security
3. Analyze various cloud computing mechanisms
4. Understanding the architecture and working of cloud computing.

Unit -I
Understanding Cloud Computing:
Origins and influences, Basic Concepts and Terminology, Goals and Benefits, Risks and Challenges.
Fundamental Concepts and Models:
Roles and Boundaries, Cloud Characteristics, Cloud Delivery Models, Cloud Deployment Models.

Unit-II
Cloud-Enabling Technology:
Broadband Networks and Internet Architecture, Data Center Technology, Virtualization Technology
CLOUD COMPUTING MECHANISMS:
Cloud Infrastructure Mechanisms: Logical Network Perimeter, Virtual Server, Cloud Storage Device, Cloud Usage Monitor, Resource Replication

Unit-III
Cloud Management Mechanisms: Remote Administration System, Resource Management System, SLA Management System, Billing Management System, Case Study Example
CLOUD COMPUTING ARCHITECTURE
Fundamental Cloud Architectures: Workload Distribution Architecture, Resource Pooling Architecture, Dynamic Scalability Architecture, Elastic Resource Capacity Architecture, Service Load Balancing Architecture, Cloud Bursting Architecture, Elastic Disk Provisioning Architecture, Redundant Storage Architecture, Case Study Example

Unit-IV
Cloud-Enabled Smart Enterprises
Introduction, Revisiting the Enterprise Journey, Service-Oriented Enterprises, Cloud Enterprises, Smart Enterprises, The Enabling Mechanisms of Smart Enterprises
Cloud-Inspired Enterprise Transformations
Introduction, The Cloud Scheme for Enterprise Success, Elucidating the Evolving Cloud Idea, Implications of the Cloud on Enterprise Strategy, Establishing a Cloud-Incorporated Business Strategy

UNIT-V Transitioning to Cloud-Centric Enterprises
The Tuning Methodology, Contract Management in the Cloud
Cloud-Instigated IT Transformations
Introduction, Explaining Cloud Infrastructures, A Briefing on Next-Generation Services, Service Infrastructures, Cloud Infrastructures, Cloud Infrastructure Solutions, Clouds for Business Continuity, The Relevance of Private Clouds, The Emergence of Enterprise Clouds

TEXT BOOKS:
1. Erl Thomas, Puttini Ricardo, Mahmood Zaigham, Cloud Computing: Concepts, Technology & Architecture 1st Edition,
2. Pethuru Raj, Cloud Enterprise Architecture, CRC Press

REFERENCES:
1. James Bond,The Enterprise Cloud, O'Reilly Media, Inc.

PREDICTIVE ANALYTICS (PROFESSIONAL ELECTIVE - IV)

M.Tech DS I Year II Sem.

L	T	P	C
3	0	0	3

Course Objectives

The course serves to advance and refine expertise on theories, approaches and techniques related to prediction and forecasting

Course Outcomes

1. Understand prediction-related principles, theories and approaches.
2. Learn model assessment and validation.
3. Understand the basics of predictive techniques and statistical approaches.
4. Analyze supervised and unsupervised algorithms

Unit I

Linear Methods for Regression and Classification: Overview of supervised learning, Linear regression models and least squares, Multiple regression, Multiple outputs, Subset selection, Ridge regression, Lasso regression, Linear Discriminant Analysis, Logistic regression, Perceptron learning algorithm.

Unit II

Model Assessment and Selection: Bias, Variance, and model complexity, Bias-variance trade off, Optimism of the training error rate, Estimate of In-sample prediction error, Effective number of parameters, Bayesian approach and BIC, Cross- validation, Bootstrap methods, conditional or expected test error.

Unit III

Additive Models, Trees, and Boosting: Generalized additive models, Regression and classification trees, Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting, Examples (Spam data, California housing, New Zealand fish, Demographic data)

Unit IV

Neural Networks (NN) , Support Vector Machines(SVM),and K-nearest Neighbor: Fitting neural networks, Back propagation, Issues in training NN, SVM for classification, Reproducing Kernels, SVM for regression, K-nearest – Neighbour classifiers(Image Scene Classification)

Unit V

Unsupervised Learning and Random forests: Association rules, Cluster analysis, Principal Components, Random forests and analysis.

TEXT BOOK:

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, Second Edition, Springer Verlag, 2009

REFERENCES:

1. C.M. Bishop –Pattern Recognition and Machine Learning, Springer,2006
2. L. Wasserman-All of statistics
3. Gareth James. Daniela Witten. Trevor Hastie Robert Tibshirani. An Introduction to Statistical Learning with Applications in R

MINING MASSIVE DATASETS (PROFESSIONAL ELECTIVE - IV)

M.Tech -DS I Year II Sem.

L	T	P	C
3	0	0	3

Prerequisites:

1. Students should be familiar with Data mining, algorithms, basic probability theory and Discrete math.

Course Objectives:

1. This course will cover practical algorithms for solving key problems in mining of massive datasets.
2. This course focuses on parallel algorithmic techniques that are used for large datasets.
3. This course will cover stream processing algorithms for data streams that arrive constantly, page ranking algorithms for web search, and online advertisement systems that are studied in detail.

Course Outcomes:

1. Handle massive data using MapReduce.
2. Develop and implement algorithms for massive data sets and methodologies in the context of data mining.
3. Understand the algorithms for extracting models and information from large datasets
4. Develop recommendation systems.
5. Gain experience in matching various algorithms for particular classes of problems.

UNIT I

Data Mining-Introduction-Definition of Data Mining-Statistical Limits on Data Mining,
MapReduce and the New Software Stack-Distributed File Systems, MapReduce, Algorithms Using MapReduce.

UNIT II

Similarity Search: Finding Similar Items**-**Applications of Near-Neighbor Search, Shingling of Documents, Similarity-Preserving Summaries of Sets, Distance Measures.
Streaming Data: Mining Data Streams-The Stream Data Model, Sampling Data in a Stream, Filtering Streams

UNIT III

Link Analysis-PageRank, Efficient Computation of PageRank, Link Spam
Frequent Itemsets-Handling Larger Datasets in Main Memory, Limited-Pass Algorithms, Counting Frequent Items in a Stream.
Clustering-The CURE Algorithm, Clustering in Non-Euclidean Spaces, Clustering for Streams and Parallelism

UNIT IV

Advertising on the Web-Issues in On-Line Advertising, On-Line Algorithms, The Matching Problem, The Adwords Problem, Adwords Implementation.
Recommendation Systems-A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering, Dimensionality Reduction, The NetFlix Challenge.

UNIT V

Mining Social-Network Graphs-Social Networks as Graphs, Clustering of Social-Network Graphs, Partitioning of Graphs, Simrank, Counting Triangles

TEXT BOOKS:

1. Jure Leskovec, Anand Rajaraman, Jeff Ullman, Mining of Massive Datasets, 3rd Edition.

REFERENCE BOOKS:

1. Jiawei Han & Micheline Kamber , Data Mining – Concepts and Techniques 3rd Edition Elsevier.
2. Margaret H Dunham, Data Mining Introductory and Advanced topics, PEA.
3. Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann.

NATURE INSPIRED COMPUTING (PROFESSIONAL ELECTIVE - IV)

M.Tech DS I Year II Sem.

L	T	P	C
3	0	0	3

Course Objectives:
Knowledge on significance of intelligence, genetic algorithms Ant Colony algorithms

Course Outcomes:
1. Familiar with Genetic algorithm and its applications.
2. Compare different Ant Colony Optimization algorithmic models.
3. Compare different Artificial Bee Colony Optimization algorithmic models.
4. Illustrate Particle swam optimization algorithm with an example.

Unit I: Models of Life and Intelligence - Fundamentals of bio-inspired models and bio-inspired computing. Evolutionary models and techniques, Swarm models and its self-organization, swarm and evolutionary algorithms. Optimization problems – single and multi-objective optimization, heuristic, meta-heuristic and hyper heuristic functions.

Unit II: Genetic algorithms - Mathematical foundation, Genetic problem solving, crossover and mutation. genetic algorithms and Markov process, applications of genetic algorithms

Unit III: Ant Colony Algorithms - Ant colony basics, hybrid ant system, ACO in combinatorial optimization, variations of ACO, case studies.

Unit IV: Particle Swarm algorithms - particles moves, particle swarm optimization, variable length PSO, applications of PSO, case studies. Artificial Bee Colony algorithms - ABC basics, ABC in optimization, multi-dimensional bee colony algorithms, applications of bee algorithms, case studies.

Unit V: Selected nature inspired techniques - Hill climbing, simulated annealing, Gaussian adaptation, Cuckoo search, Firey algorithm, SDA algorithm, bat algorithm, case studies. Other nature inspired techniques - Social spider algorithm, Cultural algorithms, Harmony search algorithm, Intelligent water drops algorithm, Artificial immune system, Flower pollination algorithm, case studies.

TEXT BOOKS:
1. Albert Y. Zomaya - "Handbook of Nature-Inspired and Innovative Computing", Springer, 2006
2. Floreano, D. and C. Mattiussi -"Bio-Inspired Artificial Intelligence: Theories methods, and Technologies" IT Press, 2008.

REFERENCES:
1. Leandro Nunes de Castro - " Fundamentals of Natural Computing, Basic Concepts, Algorithms and Applications", Chapman & Hall/ CRC, Taylor and Francis Group, 2007
2. Marco Dorrigo, Thomas Stutzle -" Ant Colony Optimization", Prentice Hall of India, New Delhi, 2005
3. Vinod Chandra S S, Anand H S - "Machine Learning: A Practitioner's Approach", Prentice Hall of India, New Delhi, 2020

BIG DATA TECHNOLOGIES LAB (LAB - III)

M.Tech DS I Year II Sem.

L	T	P	C
0	0	4	2

Course Objectives

1. The purpose of this course is to provide the students with the knowledge of Big data Analytics principles and techniques.
2. This course is also designed to give an exposure of the frontiers of Big data Analytics

Course Outcomes

1. Use Excel as an Analytical tool and visualization tool.
2. program using HADOOP and Map reduce
3. perform data analytics using ML in R.
4. Use cassandra to perform social media analytics

List of Experiments

1. Implement a simple map-reduce job that builds an inverted index on the set of input documents (Hadoop)
2. Process big data in HBase
3. Store and retrieve data in Pig
4. Perform Social media analysis using cassandra
5. Buyer event analytics using Cassandra on suitable product sales data.
6. Using Power Pivot (Excel) Perform the following on any dataset
    a. Big Data Analytics
    b. Big Data Charting
7. Use R-Project to carry out statistical analysis of big data
8. Use R-Project for data visualization of social media data

TEXT BOOKS:

1. Big Data Analytics, SeemaAcharya, Subhashini Chellappan, Wiley 2015.
2. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business, Michael Minelli, Michehe Chambers, 1st Edition, AmbigaDhiraj, Wiely CIO Series, 2013.
3. Hadoop: The Definitive Guide, Tom White, 3rd Edition, O"Reilly Media, 2012.
4. Big Data Analytics: Disruptive Technologies for Changing the Game, Arvind Sathi, 1st Edition, IBM Corporation, 2012.

REFERENCES:

1. Big Data and Business Analytics, Jay Liebowitz, Auerbach Publications, CRC press (2013)
2. Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop, Tom Plunkett, Mark Hornick, McGraw-Hill/Osborne Media (2013), Oracle press.
3. Professional Hadoop Solutions, Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Wiley, ISBN: 9788126551071, 2015.
4. Understanding Big data, Chris Eaton, Dirk deroos et al., McGraw Hill, 2012.
5. Intelligent Data Analysis, Michael Berthold, David J. Hand, Springer, 2007.
6. Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, Bill Franks, 1st Edition, Wiley and SAS Business Series, 2012.

NATURAL LANGUAGE PROCESSING LAB (LAB - IV)

M.Tech DS I Year II Sem.

L	T	P	C
0	0	4	2

Prerequisites: Data structures, finite automata and probability theory

Course Objectives:

- To Develop and explore the problems and solutions of NLP.

Course Outcomes:

1. Show sensitivity to linguistic phenomena and an ability to model them with formal grammars.
2. Able to manipulate probabilities, construct statistical models over strings and trees, and estimate parameters using supervised and unsupervised training methods.
3. Able to design, implement, and analyze NLP algorithms

List of Experiments

Implement the following using Python

1. Tokenization
2. Stemming
3. Stop word removal (a, the, are,..)
4. Word Analysis
5. Word Generation
6. Pos tagging
7. Morphology
8. chunking
9. N-Grams
10. N-Grams Smoothing

TEXT BOOKS:

1. Multilingual natural Language Processing Applications: From Theory to Practice – Daniel M. Bikel and Imed Zitouni, Pearson Publication
2. Natural Language Processing and Information Retrieval: Tanvier Siddiqui, U.S. Tiwary

REFERENCES:

1. Speech and Natural Language Processing - Daniel Jurafsky & James H Martin, Pearson Publications

EDGE ANALYTICS LAB (LAB -IV)

M.Tech DS I Year II Sem.

L	T	P	C
0	0	4	2

COURSE OBJECTIVES

1. Understand the concept of edge computing
2. Understand the Edge computing Architecture
3. Implement the edge computing in IOT
4. Understand the concept of multi-access edge computing
5. Implement edge computing in MEC

COURSE OUTCOMES

1. Identify the benefits of edge computing
2. Develop the microservices in iofog
3. Develop user defined services in the edge
4. Create use cases in IOT with edge computing
5. Develop services in MEC
6. Implement use cases in MEC

List Of Experiments:

1. Set up the Arduino IDE for ESP8266-12 module and program it to blink a LED light.
2. Installation tools to create and manage ECN's
3. Deploy micro services and writing your own microservices
4. Setup the Communication Parameters
5. Implement any two Communications protocols
6. Deploy modules to a Windows IoT Edge device
7. Create an IoT hub.
8. Register an IoT Edge device to your IoT hub.
9. Install and start the IoT Edge for Linux on Windows runtime on your device.
10. Remotely deploy a module to an IoT Edge device and send telemetry.
11. Python based basic programs using Raspberry Pi.
12. Deploy a module Manage your Azure IoT Edge device from the cloud to deploy a module that sends telemetry data to IoT Hub.
13. Publishing Data using HTTP.
14. Sensor Interfacing and Logging using MQTT.
15. File IO Example - # Example code to demonstrate writing and reading data to/from files
16. write code to turn on one of the LEDs on the board (Breadboard)

Additional Exercises on IOT Edge Analytics Applications

17. Temperature Logger
18. Home Automation

TEXT BOOKS:

1. Hands-On Edge Analytics with Azure IoT: Design and develop IoT applications with edge analytical solutions including Azure IoT Edge by Colin Dow
2. MicroPython for the Internet of ThingsA Beginner's Guide to Programming with Python on Microcontroller, Charles Bell, APress.

REFERENCE BOOKS:

1. Learn Edge Analytics - Fundamentals of Edge Analytics: Automated analytics at source using Microsoft Azure by AshishMahajan
2. Peter Waher, "Mastering Internet of Things: Design and create your own IoT applications using Raspberry Pi 3", First Edition, Packt Publishing, 2018
3. John C. Shovic, "Raspberry Pi IoT Projects: Prototyping Experiments for Makers", Packt Publishing, 2016
4. Python for Microcontrollers: Getting Started with MicroPython Paperback – 16 December 2016, by Donald Norris, McGraw-Hill Education TAB
5. Programming with MicroPython: Embedded Programming with Microcontrollers and Python, by Nicholas H. Tollervey, O'Reilly
6. R. BUYYA, S.N. SRIRAMA (2019), Fog and Edge Computing: Principles and Paradigms, Wiley-Blackwell, 2019.

ENTERPRISE CLOUD CONCEPTS LAB (LAB-IV)

M.Tech DS I Year II Sem.

L	T	P	C
0	0	4	2

Course Objectives:
Knowledge on significance of cloud computing and its fundamental concepts and models.

Course Outcomes:
1. Understand importance of cloud architecture
2. Illustrating the fundamental concepts of cloud security
3. Analyze various cloud computing mechanisms
4. Understanding the architecture and working of cloud computing.

List of Experiments:
1. Install Virtualbox/VMware Workstation with different flavors of linux or windows OS on top of windows7 or 8.
2. Install a C compiler in the virtual machine created using virtual box and execute Simple Programs
3. Install Google App Engine. Create a hello world app and other simple web applications using python/java..
4. Find a procedure to transfer the files from one virtual machine to another virtual machine.
5. Find a procedure to launch virtual machine using trystack (Online Openstack Demo Version)
6. Install Hadoop single node cluster and run simple applications like word count.

E-Resources:
1. https://www.iitk.ac.in/nt/faq/vbox.htm

2.https://www.google.com/urlsa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjqrNG0za73AhXZt1YBHZ21DWEQFnoECAMQAQ&url=http%3A%2F%2Fwww.cs.columbia.edu%2F~sedwards%2Fclasses%2F2015%2F1102-fall%2Flinuxvm.pdf&usg=AOvVaw3xZPuF5xVgk-AQnBRsTtHz

3. https://www.cloudsimtutorials.online/cloudsim/

4.https://edwardsamuel.wordpress.com/2014/10/25/tutorial-creating-openstack-instance-in-trystack/

5. https://www.edureka.co/blog/install-hadoop-single-node-hadoop-cluster

DIGITAL FORENSICS (PROFESSIONAL ELECTIVE - V)

M.Tech DS II Year I Sem.

L	T	P	C
3	0	0	3

Pre-Requisites: Cybercrime and Information Warfare, Computer Networks

Course Objectives:

1. provides an in-depth study of the rapidly changing and fascinating field of computer forensics.
2. Combines both the technical expertise and the knowledge required to investigate, detect and prevent digital crimes.
3. Knowledge on digital forensics legislations, digital crime, forensics processes and procedures, data acquisition and validation, e-discovery tools
4. E-evidence collection and preservation, investigating operating systems and file systems, network forensics, art of steganography and mobile device forensics

Course Outcomes: On completion of the course the student should be able to

1. Understand relevant legislation and codes of ethics.
2. Computer forensics and digital detective and various processes, policies and procedures.
3. E-discovery, guidelines and standards, E-evidence, tools and environment.
4. Email and web forensics and network forensics.

UNIT - I

Digital Forensics Science: Forensics science, computer forensics, and digital forensics.

Computer Crime: Criminalistics as it relates to the investigative process, analysis of cyber criminalistics area, holistic approach to cyber-forensics

UNIT - II

Cyber Crime Scene Analysis:

Discuss the various court orders etc., methods to search and seizure electronic evidence, retrieved and un-retrieved communications, Discuss the importance of understanding what court documents would be required for a criminal investigation.

UNIT - III

Evidence Management & Presentation:

Create and manage shared folders using operating system, importance of the forensic mindset, define the workload of law enforcement, Explain what the normal case would look like, Define who should be notified of a crime, parts of gathering evidence, Define and apply probable cause.

UNIT - IV

Computer Forensics: Prepare a case, Begin an investigation, Understand computer forensics workstations and software, Conduct an investigation, Complete a case, Critique a case,

Network Forensics: open-source security tools for network forensic analysis, requirements for preservation of network data.

UNIT - V

Mobile Forensics: mobile forensics techniques, mobile forensics tools.

Legal Aspects of Digital Forensics: IT Act 2000, amendment of IT Act 2008.

Recent trends in mobile forensic technique and methods to search and seizure electronic evidence

TEXT BOOKS:

1. John Sammons, The Basics of Digital Forensics, Elsevier
2. John Vacca, Computer Forensics: Computer Crime Scene Investigation, Laxmi Publications

REFERENCES

1. William Oettinger, Learn Computer Forensics: A beginner's guide to searching, analyzing, and securing digital evidence, Packt Publishing; 1st edition (30 April 2020), ISBN : 1838648178.
2. Thomas J. Holt, Adam M. Bossler, Kathryn C. Seigfried-Spellar, Cybercrime and Digital Forensics: An Introduction, Routledge.

CONVERSATIONAL AI (PROFESSIONAL ELECTIVE - V)

M.Tech DS II Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives:

1. To be familiar with the basic knowledge about conversational systems.
2. To understand the different techniques of natural language processing
3. Study the fundamental role of machine learning in building conversational systems.
4. To know the various applications of conversational systems and its future development

Course Outcomes:

1. Understand the basic technologies required for building a conversational system.
2. Learn the rule-based dialogue system
3. Involve AI in building conversational system and build advanced systems that are cognitively inclined towards human behaviour.
4. Develop a real time working conversational system for social domain that can intelligently process inputs and generate relevant replies.

UNIT-I Introducing Dialogue Systems

What's a Dialogue System? A Brief History Of Dialogue Systems, Present-Day Dialogue Systems, Modeling Conversation Dialogue Systems, Designing and Developing Dialogue Systems

UNIT-II Rule-Based Dialogue Systems: Architecture, Methods, and Tools

A Typical Dialogue Systems Architecture, Designing a Dialogue System, Tools for Developing Dialogue Systems, Rule-Based Techniques in Dialogue Systems Participating in the Alexa Prize

UNIT-III Statistical Data-Driven Dialogue Systems

Motivating the Statistical Data-Driven Approach, Dialogue Components in the Statistical Data-Driven Approach, Reinforcement Learning (RL), Representing Dialogue as a Markov Decision Process, From MDPs to POMDPs, Dialogue State Tracking, Dialogue Policy, Problems and Issues with Reinforcement Learning in POMDPs

UNIT-IV Evaluating Dialogue Systems

How to Conduct the Evaluation, Evaluating Task-Oriented Dialogue Systems, Evaluating Open-Domain Dialogue Systems, Evaluation Frameworks- PARADISE, Quality of Experience (QoE), Interaction Quality, Best Way to Evaluate Dialogue Systems.

UNIT-V End-to-End Neural Dialogue Systems

Neural Network Approaches to Dialogue Modeling, A Neural Conversational Model, Introduction to the Technology of Neural Dialogue, Retrieval-Based Response Generation, Task-Oriented Neural Dialogue Systems, Open-Domain Neural Dialogue Systems, Some Issues and Current Solutions, Dialogue Systems: Datasets, Competitions, Tasks, and Challenges.

TEXT BOOKS:

1. Michael McTear, “Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots”, Second Edition, Moran and Claypool Publishers, 2020.

REFERENCE BOOKS

1. Cathy Pearl, “Designing Voice User Interfaces: Principles of Conversational Experiences”, O’REILLY, 2016.
2. Web Services, G. Alonso, F. Casati and others, Springer.

VIDEO ANALYTICS (PROFESSIONAL ELECTIVE - V)

M.Tech DS II Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives:

1. To know the fundamental concepts of big data and analytics
2. To learn various techniques for mining data streams
3. To acquire the knowledge of extracting information from surveillance videos.
4. To learn Event Modelling for different applications.
5. To understand the models used for recognition of objects in videos.

Course Outcomes:

1. Understand the basics of video- signals and systems.
2. Able to estimate motion in a video
3. Able to detect the objects and track them
4. Recognize activity and analyze behavior
5. Evaluate face recognition technologies

UNIT I

Introduction

Multi dimensional signals and systems: signals, transforms, systems, sampling theorem. Digital Images and Video: human visual system and color, digital video, 3D video, digital-video applications, image and video quality.

UNIT II

Motion Estimation

Image formation, motion models, 2D apparent motion estimation, differential methods, matching methods, non-linear optimization methods, transform domain methods, 3D motion and structure estimation.

UNIT III

Video Analytics

Introduction- Video Basics - Fundamentals for Video Surveillance- Scene Artifacts- Object Detection and Tracking: Adaptive Background Modelling and Subtraction- Pedestrian Detection and Tracking Vehicle Detection and Tracking- Articulated Human Motion Tracking in Low-Dimensional Latent Spaces.

UNIT IV

Behavioural Analysis & Activity Recognition

Event Modelling- Behavioural Analysis- Human Activity Recognition-Complex Activity Recognition Activity modelling using 3D shape, Video summarization, shape-based activity models- Suspicious Activity Detection.

UNIT V

Human Face Recognition & Gait Analysis

Introduction: Overview of Recognition algorithms – Human Recognition using Face: Face Recognition from still images, Face Recognition from video, Evaluation of Face Recognition Technologies- Human Recognition using gait: HMM Framework for Gait Recognition, View Invariant Gait Recognition, Role of Shape and Dynamics in Gait Recognition

TEXT BOOKS:

1. A.MuratTekalp, “Digital Video Processing”, second edition, Pearson, 2015
2. Rama Chellappa, Amit K.Roy-Chowdhury, Kevin Zhou.S, “Recognition of Humans and their Activities using Video”, Morgan&Claypool Publishers, 2005.
3. Yunqian Ma, Gang Qian, “Intelligent Video Surveillance: Systems and Technology”, CRC Press (Taylor and Francis Group), 2009.

REFERENCES:

1. Richard Szeliski, “Computer Vision: Algorithms and Applications”, Springer, 2011.
2. Yao Wang, JornOstermann and Ya-Qin Zhang, “Video Processing and Communications”, Prentice Hall, 2001.

3. Thierry Bouwmans, FatihPorikli, Benjamin Höferlin and Antoine Vacavant, “Background Modeling and Foreground Detection for Video Surveillance: Traditional and Recent Approaches, Implementations, Benchmarking and Evaluation”, CRC Press, Taylor and Francis Group, 2014.
4. Md. Atiqur Rahman Ahad, “Computer Vision and Action Recognition-A Guide for Image Processing and Computer Vision Community for Action Understanding”, Atlantis Press, 2011.

STATISTICAL FOUNDATIONS FOR DATA SCIENCE (OPEN ELECTIVE)

M.Tech DS II Year I Sem.

L	T	P	C
3	0	0	3

Prerequisites: Mathematics courses of first year of study.

Course Objectives:

1. The Number Theory basic concepts useful for cryptography etc
2. The theory of Probability, and probability distributions of single and multiple random variables
3. The sampling theory and testing of hypothesis and making inferences
4. Stochastic process and Markov chains.

Course Outcomes: After learning the contents of this course, the student must be able to

1. Apply the number theory concepts to cryptography domain
2. Apply the concepts of probability and distributions to some case studies
3. Correlate the material of one unit to the material in other units
4. Resolve the potential misconceptions and hazards in each topic of study.

UNIT - I

Greatest Common Divisors and Prime Factorization: Greatest common divisors, The Euclidean algorithm, The fundamental theorem of arithmetic, Factorization of integers and the Fermat numbers Congruences: Introduction to congruences, Linear congruences, The Chinese remainder theorem, Systems of linear congruences

UNIT - II

Simple Linear Regression and Correlation: Introduction to Linear Regression, The Simple Linear Regression Model, Least Squares and the Fitted Model, Properties of the Least Squares Estimators, Inferences Concerning the Regression Coefficients, Prediction, Simple Linear Regression Case Study Random Variables and Probability Distributions: Concept of a Random Variable, Discrete Probability Distributions, Continuous Probability Distributions, Statistical Independence. Discrete Probability Distributions: Binomial Distribution, Poisson distribution.

UNIT - III

Continuous Probability Distributions: Normal Distribution, Areas under the Normal Curve, Applications of the Normal Distribution, Normal Approximation to the Binomial, Fundamental Sampling Distributions: Random Sampling, Sampling Distributions, Sampling, Distribution of Means and the Central Limit Theorem, Sampling Distribution of S2, t–Distribution, F-Distribution.

UNIT - IV

Estimation & Tests of Hypotheses: Introduction, Statistical Inference, Classical Methods of Estimation. Estimating the Mean, Standard Error of a Point Estimate, Prediction Intervals, Tolerance Limits, Estimating the Variance, Estimating a Proportion for single mean, Difference between Two Means, between Two Proportions for Two Samples and Maximum Likelihood Estimation.

UNIT - V

Stochastic Processes and Markov Chains: Introduction to Stochastic processes- Markov process. Transition Probability, Transition Probability Matrix, First order and Higher order Markov process, n step transition probabilities, Markov chain, Steady state condition, Markov analysis.

TEXT BOOKS:

1. Kenneth H. Rosen, Elementary number theory & its applications, sixth edition, Addison Wesley, ISBN 978 0-321-50031-1
2. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, Probability & Statistics for Engineers & Scientists, 9th Ed. Pearson Publishers.
3. S. D. Sharma, Operations Research, Kedarnath and Ramnath Publishers, Meerut, Delhi

REFERENCE BOOKS:

1. S C Gupta and V K Kapoor, Fundamentals of Mathematical statistics, Khanna publications
2. T.T. Soong, Fundamentals of Probability And Statistics For Engineers, John Wiley & Sons Ltd, 2004.
3. Sheldon M Ross, Probability and statistics for Engineers and scientists, Academic Press.

BIG DATA TECHNOLOGIES (OPEN ELECTIVE)

M.Tech DS II Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives

1. The purpose of this course is to provide the students with knowledge of Big data Analytics principles and techniques.
2. This course is also designed to give an exposure of the frontiers of Big data Analytics

Courses Outcomes

1. Ability to explain the foundations, definitions, and challenges of Big Data and various Analytical tools.
2. Ability to program using HADOOP and Map reduce, NOSQL
3. Ability to understand the importance of Big Data in Social Media and Mining.

Unit I

Getting an Overview of Big Data

Big Data, History of Data Management – Evolution of Big Data, Structuring Big Data, Elements of Big Data, Big Data Analytics, Careers in Big Data, Future of Big Data

Technologies for Handling Big Data

Distributed and Parallel Computing for Big Data, Introducing Hadoop, Cloud Computing and Big Data, In-Memory Computing Technology for Big Data.

Unit II

Understanding Hadoop Ecosystem

Hadoop Ecosystem, Hadoop Distributed File System, MapReduce, Hadoop YARN, Hbase, Hive, Pig and Pig Latin, Sqoop, ZooKeeper, Flume, Oozie

Understanding MapReduce Fundamentals and HBase

The MapReduce Framework, Techniques to Optimize MapReduce Jobs, Uses of MapReduce, Role of HBase in Big Data Processing

Unit III

Exploring Hive

Introducing Hive, Getting Started with Hive, Data Types in Hive, Built-In Functions in Hive, Hive DDL, Data Manipulation in Hive, Data Retrieval Queries, Using JOINS in Hive

Analyzing Data with Pig

Introducing Pig, Running Pig, Getting Started with Pig Latin, Working with Operators in Pig, Working with Functions in Pig

Unit IV

Using Oozie

Introducing Oozie, Installing and Configuring Oozie, Understanding the Oozie Workflow, Oozie Coordinator, Oozie Bundle, Oozie Parameterization with EL, Oozie Job Execution Model, Accessing Oozie, Oozie SLA

NoSQL Data Management

Introduction to NoSQL, Aggregate Data Models, Key Value Data Model, Document Databases, Relationships, Graph Databases, Schema-Less Databases, Materialized Views, Distribution Models, Sharding, MapReduce Partitioning and Combining, Composing MapReduce Calculations

Unit V

Zookeeper: Installing and Running ZooKeeper, An Example, Group Membership in ZooKeeper, Creating the Group, Joining a Group, Listing Members in a Group, The ZooKeeper Service, Data Model, Operations, Implementation, Consistency, Sessions, Building Applications with ZooKeeper, A Configuration, Service, The Resilient ZooKeeper Application, A Lock Service, More Distributed Data Structures and Protocols, ZooKeeper in Production

Sqoop: Getting Sqoop, Sqoop Connectors, A Sample Import, Generated Code, Imports: A Deeper Look, Working with Imported Data, Importing Large Objects, Performing an Export, Exports: A Deeper Look

TEXT BOOKS:

1. Big data, blackbook, DreamTech Press, 2015

2. Hadoop: The Definitive Guide, Tom White, 3rd Edition, O'Reilly Media, 2012.

REFERENCES:

1. Big Data Analytics, Seema Acharya, Subhashini Chellappan, Wiley 2015.
2. Simon Walkowiak, Big Data Analytics with R, Packt Publishing, ISBN: 9781786466457
3. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business, Michael Minelli, Michehe Chambers, 1st Edition, Ambiga Dhiraj, Wiley CIO Series, 2013.
4. Big Data Analytics: Disruptive Technologies for Changing the Game, Arvind Sathi, 1st Edition, IBM Corporation, 2012.

PREDICTIVE ANALYTICS (OPEN ELECTIVE)

M.Tech DS II Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives

The course serves to advance and refine expertise on theories, approaches and techniques related to prediction and forecasting

Course Outcomes

1. Understand prediction-related principles, theories and approaches.
2. Learn model assessment and validation.
3. Understand the basics of predictive techniques and statistical approaches.
4. Analyze supervised and unsupervised algorithms

Unit I

Linear Methods for Regression and Classification: Overview of supervised learning, Linear regression models and least squares, Multiple regression, Multiple outputs, Subset selection , Ridge regression, Lasso regression , Linear Discriminant Analysis , Logistic regression , Perceptron learning algorithm.

Unit II

Model Assessment and Selection: Bias, Variance, and model complexity, Bias-variance trade off, Optimism of the training error rate, Estimate of In-sample prediction error, Effective number of parameters, Bayesian approach and BIC, Cross- validation, Bootstrap methods, conditional or expected test error.

Unit III

Additive Models, Trees, and Boosting: Generalized additive models, Regression and classification trees , Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting ,Examples (Spam data, California housing , New Zealand fish, Demographic data)

Unit IV

Neural Networks (NN) , Support Vector Machines(SVM),and K-nearest Neighbor: Fitting neural networks, Back propagation, Issues in training NN, SVM for classification, Reproducing Kernels, SVM for regression, K-nearest –Neighbour classifiers(Image Scene Classification)

Unit V

Unsupervised Learning and Random forests: Association rules, Cluster analysis, Principal Components, Random forests and analysis.

TEXT BOOK:

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, Second Edition, Springer Verlag, 2009

REFERENCES:

1. C.M.Bishop –Pattern Recognition and Machine Learning,Springer,2006
2. L.Wasserman-All of statistics
3. Gareth James . Daniela Witten . Trevor Hastie Robert Tibshirani. An Introduction to Statistical Learning with Applications in R

GENERATIVE AI (OPEN ELECTIVE)

M.Tech DS II Year I Sem.

L	T	P	C
3	0	0	3

Course Objectives

1. To introduce the foundations, evolution, and core concepts of AI, ML, DL, NLP, and Generative AI.
2. To develop understanding of advanced neural architectures and generative models such as GANs, VAEs, and Transformers.
3. To explore Large Language Models, prompt engineering, and their real-world applications.
4. To familiarize learners with frameworks, multimodal applications, and ethical considerations in Generative AI.

Course Outcomes

1. Demonstrate knowledge of AI foundations, generative models, and advanced neural architectures.
2. Apply generative AI techniques to create solutions for text, image, video, and multimodal tasks.
3. Design, fine-tune, and optimize Large Language Models for specific applications.
4. Evaluate ethical, social, and legal implications of Generative AI deployments and propose mitigation strategies.

UNIT 1

Foundations of AI and Generative Models

Introduction and historical evolution to Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP)and Deep Learning (DL), Structure of Artificial Neural Networks (ANNs), Mathematical and computational foundations of generative modeling, Overview of generative models and their applications across various domains; Importance of Generative AI in modern applications, Transfer learning and in advancing Generative AI

UNIT 2

Advanced Neural Architectures for Generative AI

Variational Autoencoders (VAEs): principles and applications, Generative Adversarial Networks (GANs): architecture and working principles; Transformer architecture and attention mechanisms (in detail); Long Short-Term Memory Networks (LSTMs) and the limitations of traditional RNNs/LSTMs, Advanced Transformer architectures and techniques, Pre-training and transfer learning strategies for generative models

UNIT 3

Large Language Models and Prompt Engineering

Overview of Large Language Models (LLMs), GPT architecture, variants, and working principles, Pré-training and fine-tuning GPT models for applications (e.g., chatbots, text generation), Case study: GPT-based customer support chatbot, BERT architecture, pre-training objectives, and fine-tuning, Prompt Engineering: Designing effective prompts, controlling model behavior, and improving output quality, Fine-tuning language models for creative writing and chatbot development

UNIT 4

Multi-Agent Systems and Generative AI Applications

Introduction to Multi-Agent Systems (MAS),Types of agents: reactive, deliberative, hybrid, and learning agents, Multi-agent collaboration and orchestration for generative tasks, Use cases: autonomous research assistants, cooperative creative generation, distributed problem-solving, Frameworks and tools: AutoGen, CrewAI, Hugging GPT for LLM-powered multi-agent systems, Generative AI applications: Art, Creativity, Image/Video generation, Music composition, Healthcare, Finance, Real-world case studies and deployment challenges

UNIT 5

Frameworks, Multimodal Applications, and Ethics

LangChain framework: components and LLM application development, Retrieval-Augmented Generation (RAG), Embeddings, Indexing networks, and Vector databases, Generative AI across modalities: Text, Code, Image, and Video generation, Image and Video generation using GANs and VAEs, Multimodal Generative AI: integration and training strategies, Ethical considerations: bias,

fairness, trust, and responsible AI deployment, Social and legal implications of Generative AI, Risk mitigation strategies and real-world ethical case studies

TEXT BOOKS

1. Altaf Rehmani, Generative AI for Everyone: Understanding the Essentials and Applications of This Breakthrough Technology.
2. Charu C. Aggarwal, Neural Networks and Deep Learning: A Textbook.Joseph Babcock, Raghav Bali, Generative AI with Python and TensorFlow 2, 2024.

REFERENCE BOOKS

1. Josh Kalin, Generative Adversarial Networks Cookbook.
2. Jesse Sprinter, Generative AI in Software Development: Beyond the Limitations of Traditional Coding, 2024.

ONLINE REFERENCES

1. Fabian Gloeckle et al., Better & Faster Large Language Models via Multi-token Prediction, arXiv:2404.19737v1, 2024.Vaswani et al., Attention Is All You Need, NeurIPS 2017.