M.Tech. in Data Science and Engineering

Under the

Computer Science and Engineering Department

National Institute of Technology Agartala

Major Area / Department: Computer Science and Engineering

Specialization: Data Science and Engineering

Institute
National Institute of Technology Agartala
P.O. NIT Agartala
Tripura (West), India, 799046
Tel:(0381) 2346630 / 2348511
Fax: (0381) 2346360
Email: nita.director@gmail.com
http://www.nita.ac.in

Department
Computer Science and Engineering

Dr. A.P.J. Abdul Kalam Block,
NIT Agartala, Tripura (West)
India, 7990046

Vision and Mission of the Department

Vision

To be an academic leader in the areas of Computer Science and Engineering, Information Technology, and other potential areas of Computer Science with worldwide recognition.

Mission

1. Provide high quality graduate educational programs in Computer Science and Engineering.
2. Contribute significantly to the research and the discovery of new knowledge and methods in computing.
3. Offer expertise, resource, and service to the community.
4. To retain the present faculty members by providing opportunities for professional development.

Program Educational Objectives (PEO's)

PEO-1: To impart advance theoretical and practical Knowledge, enhance skills to design, test and adapt new computing technologies for attaining professional excellence and leading successful career in industries and academia.

PEO-2: To develop the ability to critically think, analyze and offer techno-commercially feasible and socially acceptable solutions to computational problems, attaining professional excellence and carrying research. & Development (R&D) effectively.

PEO-3: To work collaboratively on development of innovative systems and optimized solutions on multidisciplinary domains and exhibit high levels of professional and ethical values within organization and society globally.

PEO-4: To develop design thinking capabilities for innovation and entrepreneurship development.

Program Specific Objectives (PSOs)

PSO-1: To understand the evolutionary changes in computing, apply standard practices and strategies to promote research and development for innovative career paths and meet future challenges.

PSO-2: The ability to incorporate contemporary and evolving computational problem-solving techniques for lifelong learning support leading to higher studies and entrepreneurship development.

PSO-3: To inculcate knowledge with moral values and professional ethics, to act as a responsible citizen.

Program Outcomes (POs)

PO1: To develop the ability to apply knowledge of mathematics, engineering sciences for conducting independent research/investigation for solving practical problems.

PO2: To develop the ability to identify, formulate, conduct experiments, interpret data, synthesize information, and analyze engineering problems by writing and presenting an effective technical report/document.

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

PO4: To develop problem-solving ability to design solutions for complex engineering problems in the context of societal and environmental commitments.

PO5: To demonstrate the capability of functioning effectively as a member or team leader in software projects considering multidisciplinary environments, thus solving real-world multifaceted problems.

PO6: To develop design thinking capabilities for innovation and contribute to technological knowledge and intellectual property development.

Title of Curriculum

M.Tech. in Data Science and Engineering

Under Computer Science and Engineering Department, NIT Agartala

Objectives of the Data Science & Engineering (DSE) Specialization

The past two decades has witnessed the involvement of IT enabled services in every sector. With the proliferation of social media services, the dynamics of the World Wide Web has shifted from data consumption to a data generation environment. The social media services have enabled not only organizations but also individual users as the content providers. The Internet traffic is increasing exponentially and so is the volume of data. Applications such as social media, healthcare, e-commerce, weather forecast, traffic monitoring, etc., are producing massive amounts of data, the so-called "BIG DATA", at an unprecedented scale. This has led to a critical need for skilled professionals, popularly known as *Data Scientists*, who can mine and interpret the data. Making sense of this massive data is an exceedingly difficult challenge for scientific, technological, and industrial disciplines. Unfortunately, there is a gap between the demand and supply of data scientists and technologists due to the following reasons:

- Due to the generic nature of Undergraduate courses, they fail to address the issues in this area in a focused manner.
- There are not many postgraduate courses that focus explicitly on Data Science.

Keeping these factors in mind, the Department of CSE at NIT Agartala, proposes a two-year Masters in Technology (M.Tech.) program in Data Science and Engineering.

Major aspects of the programme,

1) Theoretical foundations: This will include the mathematical background required for the subjects.

2) Application of Theory: This will include courses where the fundamentals and advanced concepts (subjects) could be implemented.

3) Thesis/Project Work: Covering the application of the concepts learned or research,oriented work.

The programme can:

1. **Build mathematical foundations for studying DSE. (Core subjects)**
2. **Once the foundations are built, it can give options to the students to choose their domain of interest (Computer Vision, Speech, Text etc.) so that they can apply the concepts learned. (Elective subjects).**

Career opportunities:

This program would provide students an opportunity to learn both foundational and experimental components of DSE with application of Machine Learning and Deep Learning techniques. A student, on completion of this program, will be able to undertake industry careers involving innovation and problem-solving and join the industry as a **Data Scientist/Data Analyst/Data Engineer**. Along with courses that provide specialization in DSE, students will also have option to explore some applied domains such as computer vision, natural language processing, robotics, and software analysis.

Detail Syllabus	**Annexure -I**
Classrooms available Labs available	**YES** **YES**
Number of existing Faculty in the areas of Data Science and allied fields in NITA CSE Department	**8**
Duration of Program	**2 year (4 Semester)**
Total Number of Intake	**8**
Academic eligibility	**Annexure - A**

Programme Structure of M.Tech. in DSE

Semester	Subject	L	T	P	Cr.	Class Hours per week	Marks
1	1. Advanced Data Structures and Algorithms	3	1	0	4	4	100
	2. Data Mining	3	1	0	4	4	100
	3. Mathematical Foundations for Data Science	3	1	0	4	4	100
	4. Elective I * ***To be chosen from the list of electives.**	4	0	0	4	4	100
	5. Elective II* ***To be chosen from the list of electives.**	4	0	0	4	4	100
	6. Laboratory I (Advanced Data Structures and Algorithms)	0	0	2	2	3	100
	7. Laboratory II (Data Science Foundation)	0	0	2	2	3	100
	8. Seminar	0	0	1	1	2	100
	Total	**17**	**3**	**5**	**25**	**28**	**800**
2	1. Machine Learning	3	1	0	4	4	100
	2. Big Data Analytics	3	1	0	4	4	100
	3. Elective III *	4	0	0	4	4	100
	4. Elective IV *	4	0	0	4	4	100
	***To be chosen from the list of electives.**						
	5. Laboratory- I (Machine Learning Lab)	0	0	2	2	3	100
	6. Laboratory-II (Data Science Implementation)	0	0	2	2	3	100
	7. Project Preliminaries	0	0	3	3	6	100
	8. Comprehensive Viva	0	0	2	2	0	100
	Total	**14**	**2**	**9**	**25**	**28**	**800**

Semester	Subject	L	T	P	Cr.	Class Hours per week	Marks
3	Project and Thesis - I ***Students may go for industrial or inter institute collaboration, based Project work for 6 months to 1 year. The DPPC and concerned local guide may be empowered to recommend such provision.** **All existing academic rules of institute will prevail. The exact modalities may be recommended by DPPC.**	0	0	10	10	FULL	100
	Total	**0**	**0**	**10**	**10**		**100**

Semester	Subject	L	T	P	Cr.	Class Hours per week	Marks
4	Project and Thesis - II ***Students may go for industrial or inter institute collaboration, based Project work for 6 months to 1 year. The DPPC and concerned local guide may be empowered to recommend such provision.** **All existing academic rules of institute will prevail. The exact modalities may be recommended by DPPC.**	0	0	20	20	FULL	300
	Total	**0**	**0**	**20**	**20**		
Cumulative credit of the course							
Semester-I		**17**	**3**	**5**	**25**	**28**	**800**
Semester -II		**14**	**2**	**9**	**25**	**28**	**800**
Semester -III		**0**	**0**	**10**	**10**	**Full**	**100**
Semester -IV		**0**	**0**	**20**	**20**	**Full**	**300**
Total		**31**	**5**	**44**	**80**		**2000**

S. No.	List of Elective Subjects	L	T	P	Cr.	Class Hours per week	Marks
1	**Next Generation Database**	4	0	0	4	4	100
2	**Stochastic Models and Applications**	4	0	0	4	4	100
3	**Natural Language Processing**	4	0	0	4	4	100
4	**Soft Computing**	4	0	0	4	4	100
5	**Reinforcement Learning**	4	0	0	4	4	100
6	**Intrusion Detection System**	4	0	0	4	4	100
7	**Computer Vision**	4	0	0	4	4	100
8	**Information Retrieval**	4	0	0	4	4	100
9	**Recommender Systems**	4	0	0	4	4	100
10	**Deep Learning**	4	0	0	4	4	100
11	**Data Visualization**	4	0	0	4	4	100
12	**Data Science in Bioinformatics**	4	0	0	4	4	100
13	**Data Science for Decision Making**	4	0	0	4	4	100
14	**Social Network Analysis**	4	0	0	4	4	100
15	**Time Series Data Analysis**	4	0	0	4	4	100

Annexure - A

Eligibility

- As per institute norms.

Annexure- I

Detail Syllabus
Course structure for M. Tech in Data Science and Engineering, Department of CSE, NIT Agartala

Semester I

1.1 Advanced Data Structures and Algorithms	
L T P 3 ,1, 0: **4 Credits**	Prerequisites: *None*

Course Objectives:

1. The course is intended to provide the foundations of the practical implementation and usage of Algorithms and Data Structures.
2. One objective is to ensure that the student evolves into a competent programmer capable of designing and analyzing implementations of algorithms and data structures for different kinds of problems.
3. Another objective is to expose the student to the algorithm analysis techniques, to the theory of reductions, and to the classification of problems into complexity classes.

Detailed syllabus:

MODULE I
Introduction to advanced data structures, Fundamentals of the analysis of algorithms, Algorithms, Performance analysis- time complexity and space complexity, Asymptotic Notation-Big Oh, Omega and Theta notations, Complexity Analysis Examples. Data structures-Linear and nonlinear data structures, ADT concept, Linear List ADT, Recurrences: The substitution method, Recursive tree method, Masters Method, Probabilistic analysis, Amortized analysis, Randomized algorithms, Mathematical aspects, and analysis of algorithms.

MODULE II
Divide and Conquer technique, Binary search tree, AVL-trees, red-black trees, B and B+-trees, Finding the minimum and maximum, Merge sort, Quick sort, Strassen's matrix multiplication. Splay Trees, Binomial Heaps, Fibonacci Heaps, Application of k-D tree (k-dimensional tree) in range searches and nearest neighbor searches.

MODULE III
Greedy algorithms: Introduction, Knapsack problem, Job sequencing with deadlines, Minimum cost spanning trees, Kruskal's algorithm, Prim's algorithm, Optimal storage on

tapes, Optimal merge pattern, Subset cover problem, Container loading or Bin packing problem.

MODULE IV
Dynamic algorithms: Introduction Dynamic algorithms, All pair shortest path, 0/1 knapsack, Travelling salesman problem, Coin Changing Problem, Matrix Chain Multiplication, Flow shop scheduling, Optimal binary search tree (OBST), Analysis of All problems, Introduction to NP-Hard And NP- Complete Problems
More algorithms: Dynamic programming, graph algorithms: DFS, BFS, topological sorting, shortest path algorithms, network flow problems.

MODULE IV
String Matching: The naïve string-matching algorithm, Rabin Karp algorithm, KnuthMorrisPratt algorithm (KMP), longest common subsequence (LCS), Fractional cascading, suffix trees, geometric algorithms.

References:

1. Cormen, Leiserson, Rivest and Stein, Introduction to algorithms (Main textbook)
2. Kleinberg and Tardos , Algorithm Design
3. Mark Weiss, Data structures and algorithm analysis in C++ (Java)
4. Aho, Hopcroft and Ullman, Data structures and algorithms
5. S. Sahni, Data Structures, Algorithms, and Applications in C++, Silicon Press

Course Outcome (CO):

Course Outcome No.	**Course Outcome**
CO1	Basic ability to analyze algorithms and to determine algorithm correctness and timeEfficiency class.
CO2	Master a variety of advanced abstract data type (ADT) and data structures and theirimplementations.
CO3	Master different algorithm design techniques (brute-force, divide and conquer,greedy, etc
CO4	Ability to apply and implement learned algorithm design techniques and datastructures to solve problem.
CO5	Ability to crawl information and explain different types of search algorithms.

CO-PO Mapping:

Levels: 1: Slight (LOW) 2: Moderate (MEDIUM) 3: Substantial (HIGH) and for NO CORELATION “--”

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	1	1	1	3	1	3
CO2	2	3	3	1	2	3
CO3	2	2	2	2	3	2
CO4	3	3	1	3	3	2
CO5	1	1	2	2	1	0
Total	9	10	9	11	10	10
Average Attainment	2.25	2.5	2.25	2.2	2.5	2.5
Eq. Average Attainment	2	2	2	2	3	3

1.2. Data Mining	
L T P 3, 1, 0 : **4 Credits**	Prerequisites: *None*

Course Objective:

1. To understand Data Mining in Knowledge discovery process, and its applications.
2. To understand different data attribute types and apply different data preprocessing techniques.
3. To understand how to identify association among data objects by learning various association mining algorithms.
4. To understand the various classification techniques, their applications in different domains.
5. To understand the various clustering techniques, their applications in different domains.
6. To learn various data visualization techniques for data analysis.

Detailed syllabus:

MODULE I
Introduction: Data Mining, Motivation, Application, Data Mining—On What Kind of Data? Data Mining Functionalities, Data Mining Task Primitives, Major Issues in Data Mining.
Data pre-processing: Attribute types, Similarity & Dissimilarity measures.

MODULE II
Data Preprocessing: Data Cleaning, Data Integration, Data Reduction, Data Transformation & Discretization.

MODULE III
Mining Frequent Patterns: Basic Algorithms, Association Rule Mining, Apriori Algorithm, FP tree growth Algorithm, Advanced Pattern Mining Techniques.

MODULE IV
Classification Techniques: Decision Tree, Bayes Classification, Bayesian Belief Networks, Support Vector Machines, Classification Evaluation Techniques, Classification Accuracy improvement Techniques.

MODULE V
Clustering Techniques: Partitioning algorithms, Hierarchical algorithms, Density-Based algorithms, Grid-Based algorithms, Evaluation of Clustering. Outlier Detection Techniques.

MODULE VI
Applications and Trends in Data Mining: Applications, Advanced Techniques, Web Mining, Web Content Mining, Structure Mining.

Text Books:

1. J. Han and M. Kamber. Data Mining: Concepts and Techniques. 3rd Edition, Morgan Kaufman. Pang Ning Tan, Introduction to Data Mining, 2nd Edition, Pearson.
2. M. H. Dunham. Data Mining: Introductory and Advanced Topics. Pearson Education. Roiger & Geatz, Data Mining, Pearson Education
3. A.K.Pujari, Data Mining, University Press

References Books:

1. Charu C. Aggarwal, Data Mining: The Textbook, Springer.
2. I. H. Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
3. D. Hand, H. Mannila and P. Smyth. Principles of Data Mining. Prentice, Hall.

Course Outcome (CO):

CO Number	Course Outcome
CO1	Students will be able to interpret the contribution of data mining in knowledge discovery process.
CO2	Students will be able to identify different data attribute types and apply different data preprocessing techniques.
CO3	Students will be able to apply the link analysis and frequent item-set algorithms to identify the entities on the real-world data.
CO4	Students will be able to apply the various classification and clustering algorithms for supervised and unsupervised learning problems.
CO5	Students will be able to apply various data visualization techniques in-depth data analysis.
CO6	Students will be able to apply the advanced data mining techniques and use the popular data mining tools.

CO-PO Mapping:

Levels: 1: Slight (LOW) 2: Moderate (MEDIUM) 3: Substantial (HIGH) and for NO CORELATION “--”

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	2	2	1	1	--
CO2	2	2	2	--	--	--
CO3	3	2	3	1	--	--
CO4	3	3	3	1	2	1
CO5	3	3	3	1	2	1
CO6	3	3	3	1	3	2
Total	16	15	16	5	8	4
Average Attainment	2.7	2.5	2.7	0.8	1.3	0.67
Equivalent Average Attainment	3	3	3	1	1	1

1.3. Mathematical Foundations for Data Science	
L T P 3 , 1 , 0 : **4 Credits**	Prerequisites: *None*

Course Objectives:

1. To introduce students to the various Mathematical concepts to be used in ML and DS
2. Learn the concepts of probability and Statistics.
3. Learn how to pose optimization problems.
4. Learn how to solve problems by using different algorithms.

Detailed syllabus:

MODULE I
Basics of Linear Algebra: Representation of vectors; Linear dependence and independence; vector space and subspaces (definition, examples, and concepts of basis); linear transformations; range and null space; matrices associated with linear transformations; special matrices; eigenvalues and eigenvectors with applications to data problems; Least square and minimum normed solutions.

MODULE II
Matrices in Machine Learning Algorithms: projection transformation; orthogonal decomposition; singular value decomposition; principal component analysis and linear discriminant analysis.

MODULE III
Gradient Calculus: Basic concepts of calculus: partial derivatives, gradient, directional derivatives, Jacobian, Hessian matrix.

MODULE IV
Optimization: Convex sets, convex function, and their properties, Unconstrained and Constrained Optimization, Numerical Optimization Techniques for Constrained/Unconstrained Optimization, Derivative-Free methods (Golden Section, Fibonacci Search Method, Bisecting Method), Methods using Derivatives (Newton's Method, Steepest Descent Method), Penalty Function Methods for Constrained Optimization.

MODULE V
Probability: Basic concepts of probability, conditional probability, total probability, independent events, Bayes' theorem, random variable, Moments, moment generating functions, some useful distributions, Joint distribution, conditional distribution, transformation of random variables, covariance, correlation.

MODULE VI
Statistics: Random sample, sampling techniques, statistics, sampling distributions, mixture models.

Text Books:

1. M. P. Deisenroth, A. A. Faisal, C. S. Ong, Mathematics for Machine Learning, Cambridge University Press (1st edition)
2. S. Axler, Linear Algebra Done Right. Springer International Publishing (3rd edition)
3. J. Nocedal and S. J. Wright, Numerical Optimization. New York: Springer Science+Business Media
4. E. Kreyszig, Advanced Engineering Mathematics, John Wiley and Sons, Inc., U.K. (10th Edition)
5. R. A. Johnson, I. Miller, and J. E.Freund, "Miller & Freund's Probability and Statistics for Engineers", Prentice Hall PTR, (8th edition)
6. C. Mohan and K. Deep: "Optimization Techniques", New Age Publishers, New Delhi.

Course Outcomes:

CO-No.	Course Outcome
1	To acquire knowledge on various Mathematical concepts to be used in Machine Learning and Data Science.
2	To apply the concepts of probability and Statistics.
3	To solve the various problems using optimization problems.
4	To solve various problems on data science using different algorithms.

CO-PO Mapping (Rate: scale of 1 to 3)

Course Outcome	PO-1	PO-2	PO-3	PO-4	PO-5	PO-6
CO-1	3	3	2	2	1	--
CO-2	3	2	2	2	1	--
CO-3	2	2	2	2	1	1
CO-4	2	2	2	2	1	1
Total	10	9	8	8	4	2
Average	2.5	2.25	2	2	1	0.5
Attainment	3	2	2	2	1	1

Where Levels: 1: Slight (LOW) 2: Moderate (MEDIUM) 3: Substantial (HIGH) and for NO CORELATION "--"

Laboratory of 1st Semester

1.6. Laboratory I (Advanced Data Structures and Algorithms)	
L T P 0, 0, 2 : **2 Credits**	Prerequisites: *None*

Course Outcomes:

1. Describe how arrays, records, linked structures, stacks, queues, trees, and graphs are represented in memory and used by algorithms [ABET (a, b, c, i)].
2. Describe common applications for arrays, records, linked structures, stacks, queues, trees, and graphs [ABET (a, b, c)] .
3. Write programs that use arrays, records, linked structures, stacks, queues, trees, and graphs [ABET (a, c)]
4. Demonstrate different methods for traversing trees [ABET (a)].

Programme Outcomes:

1. Identify, formulate, and analyze complex engineering problems reaching substantiated conclusions using first principles engineering sciences.

Experiment 1 (Arrays, Linked List, Stacks, Queues, Binary Trees)

I. WAP to implement a 3 stacks of size 'm' in an array of size 'n' with all the basic operations such as IsEmpty(i), Push(i), Pop(i), IsFull(i) where 'i' denotes the stack number (1,2,3), m ≅ n/3. Stacks are not overlapping each other. Leftmost stack facing
the left direction and other two stacks are facing in the right direction.

II. WAP to implement 2 overlapping queues in an array of size 'N'. There are facing in opposite direction to each other. Give IsEmpty(i), Insert(i), Delete(i) and IsFull(i) routines for i^{th} queue

III. WAP to implement Stack ADT using Linked list with the basic operations as Create(), Is Empty(), Push(), Pop(), IsFull() with appropriate prototype to a functions.

IV. WAP to implement Queue ADT using Linked list with the basic functions of Create(), IsEmpty(), Insert(), Delete() and IsFull() with suitable prototype to a functions

Experiment 2 (Sorting & Searching Techniques)

Experiment 3 (Hashing)

I. WAP to store k keys into an array of size n at the location computed using a hash function, loc = key % n, where k<=n and k takes values from [1 to m], m>n. To handle the collisions use the following collision resolution techniques, a. Linear, Quadratic, Random probing, Double hashing/rehashing, Chaining.

Experiment 4 (BST and Threaded Trees) Experiment 5 (AVL Trees and Red,Black Trees) Experiment 6 (B,Trees)
Experiment 7 (Min,Max Heaps, Binomial Heaps and Fibonacci Heaps)
Experiment 8 (Disjoint Sets) Experiment 9 (Graphs Algorithms)

No. of Course Outcome (CO)	Course Outcome
CO1	Describe how arrays, records, linked structures, stacks, queues, trees, and graphs are represented in memory and used by algorithms [ABET (a, b, c,i)].
CO2	Describe common applications for arrays, records, linked structures, stacks,queues, trees, and graphs [ABET (a, b, c) .
CO3	Write programs that use arrays, records, linked structures, stacks, queues,trees, and graphs[ABET (a, c)]
CO4	Demonstrate different methods for traversing trees [ABET (a)].

Experiment 10 (String Matching)

Course Outcomes (CO):

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	1	3	1	3
CO2	2	1	3	1	2	3
CO3	2	2	2	2	2	2
CO4	3	3	3	3	3	1
Total	9	7	9	9	8	9
Average	2.25	1.75	2.25	2.25	2	2.25
Eq. Avg. Attainment	2	2	2	2	2	2

1.7. Laboratory II (Data Science Foundations)	
L T P 0, 0, 2 : **2 Credits**	Prerequisites: *None*

Course Objective:

1. Become familiar with basic Python libraries such as NumPy, Pandas, Matplotlib, Scikit-Learn.
2. To understand and investigate the statistical nature of the data.
3. To understand the importance of data preprocessing techniques.
4. To implement basic data mining algorithms.

List of Experiments

MODULE I

1. Study of Python data types and functions.
2. Study of Python NumPy library to create multi-dimensional arrays and find its shape and dimension, create a matrix full of zeros and ones, reshape and flatten data in the array, append data vertically and horizontally, apply indexing and slicing on array.
3. To implement dot and matrix product of two arrays, compute the Eigen values of a matrix, solve a linear matrix equation, Compute the multiplicative inverse of a matrix, Compute the rank of a matrix, and compute the determinant of an array.

MODULE II

1. Study of Python Pandas library.
2. Loading data from CSV and Excel file, Compute the basic statistics of given data - shape, no. of columns, mean, standard deviation.
3. Visualization of the data distribution.

MODULE III

1. To understand the problem of data preprocessing.
2. Load data, describe the given data and identify missing, outlier data items, find correlation among all attributes, visualize correlation matrix.
3. Apply data transformation techniques- data discretization (binning etc.), data normalization ((MinMaxScaler or MaxAbsScaler).

MODULE IV

1. Implementation of association rule mining algorithms.
2. Implementation of frequent pattern mining algorithms.

Course Outcome (CO):

CO Number	Course Outcome
CO1	Demonstrate fundamental understanding of the important Python libraries required for Data Science.
CO2	Understand the statistical nature of data using measures of central tendency, measures of dispersion.
CO3	Understand and implement the concepts of data preprocessing techniques.
CO4	Understand and implement the fundamental data mining algorithms.

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	2	1	1	--
CO2	2	2	3	1	1	--
CO3	2	2	2	2	1	--
CO4	3	3	3	2	1	1
Total	9	8	10	6	4	1
Average	2.25	2	2.5	1.5	1	0.5
Eq. Avg. Attainment	2	2	3	2	1	1

Semester II

2.1. Machine Learning	
L T P 3, 1, 0: **4 Credits**	Prerequisites: *None*

Course Objectives:

1. To recognize the characteristics of machine learning that makes it useful to solve real-world problems.
2. To understand the appropriate implementation of supervised, semi supervised and unsupervised learning techniques in real-world applications.
3. To choose a suitable machine learning model, implement, and examine the performance of the chosen model for a given real world problem.
4. To understand cutting edge technologies related to machine learning applications.

Detailed Syllabus:

MODULE I
Introduction: Definition of learning systems. Goals and applications of machine learning. Aspects of developing a learning system: training data, concept representation, function approximation. The concept learning task. Concept learning as search through a hypothesis space. General-to-specific ordering of hypotheses. Finding maximally specific hypotheses. Version spaces and the candidate elimination algorithm. Learning conjunctive concepts. The importance of inductive bias.

MODULE II
Supervised Learning: Classification vs. Regression, Linear and Logistic Regression, Gradient Descent, Support Vector Machines, Kernels, Decision Trees, ML and MAP Estimates, K-Nearest Neighbor, Naive Bayes, Introduction to Bayesian Networks, Artificial Neural Networks.

MODULE III
Unsupervised Learning: Partitioning based methods, Hierarchical methods, Density based methods, Gaussian Mixture Models, Learning with Partially Observable Data (EM). Dimensionality Reduction and Principal Component Analysis.

MODULE IV
Optimization Techniques: Bias-Variance tradeoff, Regularization, Evaluation techniques for supervised and unsupervised learning.

MODULE V

Other Learning techniques: Semi-supervised Learning, Active Learning, Reinforcement Learning.

MODULE VI

Recommender System: Recommender system functions, understanding ratings, Applications of recommendation systems, Issues with recommender system, Collaborative Filtering, Content based recommendation.

Textbooks:

1. T. Mitchell, Machine Learning, McGrawHill.
2. Ethem Alpaydın, Introduction to Machine Learning 3rd Edition, MIT Press
3. Kevin Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012

References:

1. Marc Peter Deisenroth, A. Aldo Faisal and Cheng Soon Ong, Mathematics for Machine Learning, Cambridge University Press, 2020.
2. Shwartz and David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press.
3. C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
4. Andrew Ng, Machine Learning Yearning.
5. Other online material.

Course Outcomes (CO):

CO No.	**Course Outcome**
CO1	Students will be able to understand the mathematics and engineering sciences behind functioning of machine learning.
CO2	Students will be able to analyze the given dataset and data attributes for designing a machine learning based solution.
CO3	Students will be able to identify different machine learning approaches, optimization techniques, and apply them on different problem domains.
CO4	Students will be able to design and deploy machine learning solutions for real-world applications with popular machine learning tools.

CO-PO Mapping:

Course Outcome	PO-1	PO-2	PO-3	PO-4	PO-5	PO-6
CO-1	2	2	2	1	1	--
CO-2	2	2	1	2	2	--
CO-3	3	3	3	3	2	1
CO-4	3	3	2	1	2	2
Total	10	10	8	7	7	3
Average	2.5	2.5	2	1.75	1.75	0.75
Attainment	3	3	2	2	2	1

2.2. Big Data Analytics	
L T P 3 , 1 , 0 : **4 Credits**	Prerequisites: *None*

Course Objectives:

1. To explore the fundamental concepts of big data analytics using intelligent techniques.
2. To learn to use various techniques for mining data stream.
3. To understand the applications using Map Reduce Concepts.
4. To understand the programming tools and frameworks in Hadoop distributed system.
5. To acquire the knowledge of different big data issues.

Detailed syllabus:

MODULE I
Introduction to big data: Introduction to Big Data Platform; Challenges of Conventional Systems, Intelligent data analysis; Nature of Data, Analytic Processes and Tools, Analysis vs Reporting, the four dimensions of Big Data: volume, velocity, variety, veracity, Drivers for Big Data, Introducing the Storage, Query Stack, Revisit useful technologies and concepts, Real-time Big Data Analytics.

MODULE II
Mining data streams: Introduction to Streams Concepts; Stream Data Model and Architecture, Stream Computing, Sampling Data in a Stream; Filtering Streams; Counting Distinct Elements in a Stream; Estimating Moments; Counting Oneness in a Window; Decaying Window, Real time Analytics Platform (RTAP) Applications, Case Studies, Real Time Sentiment Analysis, Stock Market Predictions.

MODULE III
Distributed File Systems: Hadoop Distributed File System History of Hadoop- the Hadoop Distributed File System; Components of Hadoop Analyzing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS Basics- Developing a Map Reduce Application-How Map Reduce Works-Anatomy of a Map Reduce Job Run-Failures-Job Scheduling-Shuffle and Sort; Task execution, Map Reduce Types and Formats- Map Reduce Features Hadoop environment. Data Consistency.

MODULE IV
Overview of Spark Ecosystem, Understanding Spark Cluster Modes on YARN, RDDs (Resilient Distributed Datasets), General RDD Operations: Transformations & Actions, Common Spark

Use Cases, Data Frames and Spark SQL, Analyzing Data with Pig, NoSQL and HBase

MODULE V

Scalable Algorithms: Mining large graphs, with focus on social networks and web graphs. Centrality, similarity, a 11-distances sketches, community detection, link analysis, spectral techniques. Map-reduce, Pig Latin, and NoSQL using MongoDB, Algorithms for detecting similar items, Recommendation systems, Data stream analysis algorithms, clustering algorithms, Detecting frequent items.

MODULE VI

Frameworks and Big Data Issues: Applications on Big Data Using Pig and Hive; Data processing operators in Pig; Hive services; HiveQL; Querying Data in Hive, fundamentals of HBase and ZooKeeper, IBM InfoSphere BigInsights and Streams. Privacy, Visualization, Compliance and Security, Structured vs Unstructured Data.

Text Books:

1. Ohlhorst, Frank J. Big data analytics: turning big data into big money. Vol. 65. John Wiley & Sons, 2012.
2. Russom, Philip. "Big data analytics." TDWI best practices report, fourth quarter 19, no. 4 (2011): 1-34.
3. Marr, Bernard. Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance. John Wiley & Sons, 2015.
4. LaValle, Steve, Eric Lesser, Rebecca Shockley, Michael S. Hopkins, and Nina Kruschwitz. "Big data, analytics and the path from insights to value." MIT sloan management review 52, no. 2 (2011): 21-32.
5. Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. Mining of massive data sets. Cambridge university press, 2020.
6. Michael Berthold, David J. Hand, “Intelligent Data Analysis”, Springer, 2007.
7. Tom White “Hadoop: The Definitive Guide” Third Edition, O’reilly Media, 2012.
8. Chris Eaton, Dirk De Roos, Tom Deutsch, George Lapis, Paul Zikopoulos, “Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data”, McGrawHill Publishing, 2012.
9. Arshdeep Bahga, Vijay Madisetti, “Big Data Science & Analytics: A Hands On Approach “,VPT, 2016

Course Outcome (CO):

CO No.	Course Outcome
CO1	Acquire the fundamental concepts of big data analytics using Intelligent techniques.
CO2	To learn how to use various techniques for mining data stream.
CO3	Map Reduce Concepts implementation in Big data problem.
CO4	Acquire the knowledge of programming tools and frameworks in Hadoop distributed system.
CO5	To explore different issues in big data domain.

CO-PO Mapping:

Course Outcome	PO-1	PO-2	PO-3	PO-4	PO-5	PO-6
CO-1	1	1	1	1	1	--
CO-2	1	2	2	1	1	--
CO-3	2	2	2	1	1	1
CO-4	3	2	2	1	2	1
CO-5	3	3	2	2	2	1
Total	10	10	9	6	7	3
Average Attainment	2	2	1.8	1.2	1.4	0.6
Eq. Average Attainment	2	2	2	1	1	1

Laboratory of 2nd Semester

2.6 Laboratory-I (Machine Learning Lab)	
L T P 0,0,3: 2Credits	Prerequisites: None

Course Objective:

1. To recognize data attribute types and data preprocessing techniques.
2. To understand and apply supervised, unsupervised, and other learning techniques.
3. To understand and apply machine learning optimization techniques.
4. To understand and apply various machine learning algorithm performance evaluation techniques.
5. To choose a suitable machine learning model, implement, and examine the performance of the chosen model for a given real world problem.
6. To understand cutting edge technologies related to machine learning applications.

Detailed Syllabus:

MODULE I
Data preprocessing: Introduction to NumPy, Pandas, matplotlib, Scikit-learn.

MODULE II
Supervised Learning: Implementation of Linear and logistic regression, Naïve bayes, Decision Tree, Support Vector Machines, Neural Networks.

MODULE III
Unsupervised Learning: Implementation of k-means, Agglomerative, DBSCAN, Dimensionality Reduction and Principal Component Analysis.

MODULE IV
Optimization Techniques: Bias-Variance tradeoff, Cross-validation, Regularization, Precision, Recall and F-measure.

MODULE V
Other Learning techniques: Implementation of Reinforcement Learning, Recommender Systems, Anomaly Detection.

MODULE VI
Applications of Machine Learning: Texts, Image, Time-series data.

Textbooks:

1. Andreas C. Müller and Sarah Guido, Introduction to Machine Learning with Python: A Guide for Data Scientists, O'Reilly.
2. Other online material.

Course Outcome (CO):

Course Outcome No.	Course Outcome
CO1	Students will be able to understand the mathematics and engineeringsciences behind functioning of machine learning.
CO2	Students will be able to analyze the given dataset and data attributesfor designing a machine learning-based solution.
CO3	Students will be able to identify different machine learning approaches, optimization techniques, and apply them on different problem domains.
CO4	Students will be able to design and deploy machine learning solutions for real-world applications with popular machine learning tools.

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	2	2	1	1	--
CO2	2	2	1	2	2	--
CO3	3	3	3	3	2	1
CO4	3	3	2	1	2	2
Total	10	10	8	7	7	3
Average Attainment	2.5	2.5	2	1.75	1.75	0.75
Eq. Average Attainment	3	3	2	2	2	1

2.7 Laboratory-II (Data Science Implementation)	
L T P 0,0,3:2Credits	Prerequisites: None

Course Objective:

1. To understand and implement dimensionality reduction techniques.
2. To understand and implement vectorization of data.
3. To understand and implement classification algorithms.
4. To understand and implement clustering algorithms.

List of Experiments

MODULE I

Study and implementation of various dimensionality reduction techniques.

1. Implementation of feature selection techniques: Missing values, Low Variance Filter, High Correlation Filter, Random Forest, Backward Feature Elimination, Forward Feature Selection.
2. Implementation of Factor Analysis techniques: Principal Component Analysis (PCA), Independent Component Analysis, Linear Discriminant Analysis, Singular Valued Decomposition (SVD).
3. Implementation of Projection techniques: Isometric mapping (ISOMAP), t-distributed Stochastic Neighbor Embedding (t-SNE), Uniform Manifold Approximation and Projection (UMAP).

MODULE II

Study and implementation of various Vectorization techniques.

1. Implementation of term-frequency-inverse-document-frequency (tf-idf).
2. Implementation of Word2Vec embeddings.
3. Implementation of GloVe embeddings.
4. Implementation of FastText embeddings.
5. Other vectorization techniques.

MODULE III

Study and implementation of classification techniques.

1. Implementation of Decision Tree.
2. Implementation of Naïve Bayes.

MODULE IV

Study and implementation of clustering techniques.

1. Implementation of partitioning-based clustering algorithms: k-means.

2. Implementation of hierarchical clustering algorithms: Agglomerative, Divisive.
3. Implementation of density-based clustering algorithms: DBSCAN, HDBSCAN.

Course Outcome (CO):

CO Number	Course Outcome
CO1	Demonstrate fundamental understanding of the important dimensionality reduction techniques required for Data Science.
CO2	Understand and to be able to implement the various vectorization techniques.
CO3	Understand and to be able to implement the concepts of classification techniques.
CO4	Understand and to be able to implement the concepts of clustering techniques.

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	2	2	2	1	0
CO2	2	2	3	2	1	1
CO3	2	3	3	3	1	0
CO4	3	3	3	3	1	1
Total	9	10	11	10	4	1
Average	2.25	2.5	2.5	2.5	1	0.75
Eq. Avg. Attainment	2	3	3	3	1	1

Detailed Syllabus of Electives

1. Next Generation Database	
L T P 3 , 0 , 0 : **4 Credits**	Prerequisites: *None*

Course Objectives:

1. Understand the Database and Big data revolution.
2. To learn about NoSQL databases and their concepts.
3. To comprehend and apply columnar and distributed database patterns.
4. To learn how to use various data models for a wide range of databases.

Detailed syllabus:

MODULE I

Database Revolutions -- System Architecture, Relational Database, Database Design Data Storage, Transaction Management, Data warehouse and Data Mining, Information Retrieval.

MODULE II

Big Data Revolution -- CAP Theorem, Birth of NoSQL, Document Database - XML Databases, JSON Document Databases, Graph Databases.

MODULE III

Column Databases -- Data Warehousing Schemes, Columnar Alternative, Sybase IQ, C Store and Vertica, Column Database Architectures, SSD and In-Memory Databases, In Memory Databases, Berkeley Analytics Data Stack and Spark.

MODULE IV

Distributed Database Patterns -- Distributed Relational Databases, Non-relational Distributed Databases, MongoDB, Sharing and Replication, HBase, Cassandra Consistency Models, Types of Consistency, Consistency MongoDB, HBase Consistency, Cassandra Consistency.

MODULE V

Data Models and Storage -- SQL, NoSQL APIs, Return SQL, Advance Databases—PostgreSQL, Riak, CouchDB, NEO4J, Redis, Future Databases, Revolution Revisited Counter revolutionaries, Oracle HQ, Other Convergent Databases, Disruptive Database Technologies.

Text Books:

1. Abraham Silberschatz, Henry F. Korth, S. Sudarshan, “Database System Concepts”, Sixth Edition, McGrawHill.
2. Guy Harrison, “Next Generation Databases”, Apress, 2015.

3. Eric Redmond, Jim R Wilson, “Seven Databases in Seven Weeks”, LLC. 2012.
4. Dan Sullivan, “NoSQL for Mere Mortals”, Addison-Wesley, 2015.
5. Adam Fowler, “NoSQL for Dummies “, John Wiley & Sons, 2015.

Course Outcome (CO):

CO Number	Course Outcome
CO1	Analyze the characteristics, architecture of database and big data.
CO2	Formulate solutions to a broad range of query problems using
CO3	NoSQL concepts and relational algebra Design the big data problems using columnar and distributed database patterns
CO4	Implement the isolation property using serializabilty and concurrency control techniques

CO-PO Mapping:

Levels: 1: Slight (LOW) 2: Moderate (MEDIUM) 3: Substantial (HIGH) and for NO CORELATION “--”

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	1	1	2	2	1	--
CO2	2	2	2	2	1	--
CO3	2	2	2	2	1	1
CO4	2	3	3	3	2	1
Total	7	8	9	9	5	2
Average Attainment	1.75	2	2.25	2.25	1.25	0.5
Equivalent Average Attainment	2	2	2	2	1	1

2. Stochastic Models and Applications	
L T P 4 , 0 , 0 : **4 Credits**	Prerequisites: *None*

Course Objectives:

1. Understand the need for system models that capture random behavior to assess the risk of undesirable outcomes.
2. Be able to model several important industrial and service systems and analyze those models to improve system performance.
3. Be able to construct algorithmic solution strategies to explore system models that have been developed.

Detailed Syllabus:

MODULE I

Introductory Probability: Defining Random Variables (RVs) Events, Measurability, Independence Sample Spaces, Events, Measures, Probability, Independence, Conditional probability, Bayes' theorem Random Variables. RVs: Bernoulli, Binomial, Geometric, Poisson, Uniform, Exponential, Normal, Lognormal, Expectations, Moments and Moment generating functions Random Vectors. Random Vectors: Joint and Marginal distributions, Dependence, Covariance, Copulas, Transformations of random vectors, Order statistics.

MODULE II

Intermediate Probability: Manipulating RVs Conditioning RVs. Conditional Distribution of a RV, Computing probabilities and expectations by conditioning, RVs Distributions.

Inequalities: Markov, Chebyshev, Jensen, Holder, Convergence of RVs: Weak and Strong laws, Central limit theorem, Distributions of extreme.

MODULE III

Stochastic Processes: Indexing RVs Markov Chains, Markovian property and Transition probabilities, Irreducibility and Steady, State probabilities.

Generic Applications: Hidden Markov Chains Exponential Distribution and Poisson Process, Construction of Poisson Process from Exponential Distribution, Thinning and Conditional Arrival Times, Service Applications: Waiting Times Normal Distribution and Brownian Process, Construction of Brownian Process from Normal Distribution, Hitting Times and Maximum Values, Finance Applications: Option Pricing and Arbitrage Theorem

References:

1. Introduction to Stochastic Processes. S.M. Ross. Adventures in Stochastic Processes. S. Resnick. Birkhauser
2. Comparison Methods for Stochastic Models and Risks. A. Muller and D. Stoyan. John Wiley & Sons Mathematical Theoy of Reliability. R.E. Barlow and F. Proschan.

Course Outcome:

CO Number	Course Outcome
CO1	Students would acquire a rigorous understanding of basic concepts inprobability theory.
CO2	Learn some important concepts concerning multiple random variables such as Bayes rule for random variables, conditional expectation and its uses etc.
CO3	Explain and work on stochastic processes, including Markov Chains and Poisson Processes.

CO-PO Mapping: (Rate on a scale of 1 to 3)

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	1	1	3	1	3	1
CO2	2	3	3	2	3	2
CO3	3	3	2	2	2	2
Total	6	7	8	5	8	5
Average	2	2.33	2.67	1.67	2.67	1.67
Eq. Avg. Attainment	2	2	2	1	2	1

3. Natural Language Processing	
L T P 4 , 0 , 0 : **4 Credits**	Prerequisites: *None*

Course Objective:

1. Teach students the leading trends and systems in natural language processing.
2. Make them understand the concepts of morphology, syntax, semantics, and pragmatics of the language and that they can give the appropriate examples that will illustrate the mentioned concepts in the syllabus.
3. Teach them to recognize the significance of pragmatics for natural language understanding.
4. Enable students to be capable to describe the application based on natural language processing and to show the points of syntactic, semantic, and pragmatic processing.

Detailed syllabus:

MODULE I
Sound: Biology of Speech Processing; Place and Manner of Articulation; Word Boundary Detection; Argmax based computations; HMM and Speech Recognition.

MODULE II
Words and Word Forms: Morphology fundamentals; Morphological Diversity of Indian Languages; Morphology Paradigms; Finite State Machine Based Morphology; Automatic Morphology Learning; Shallow Parsing; Named Entities; Maximum Entropy Models; Random Fields.

MODULE III
Structures: Theories of Parsing, Parsing Algorithms; Robust and Scalable Parsing on Noisy Text as in Web documents; Hybrid of Rule Based and Probabilistic Parsing; Scope Ambiguity and Attachment Ambiguity resolution.

MODULE IV
Meaning and pragmatics: Lexical Knowledge Networks, Wordnet Theory; Indian Language Wordnets and Multilingual Dictionaries; Semantic Roles; Word Sense Disambiguation; WSD and Multilinguality; Metaphors; Coreferences. Discourse, Dialogue and Conversational agents, Natural Language Generation, Machine Translation.

MODULE V
Web 2.0 Applications: Sentiment Analysis; Text Entailment; Robust and Scalable Machine Translation; Question Answering in Multilingual Setting; Cross Lingual Information Retrieval

(CLIR).

References:

1. Speech and Language Processing by Daniel Jurafsky, James H. Martin, Second Edition, Prentice Hall
2. James Allen, "Natural Language Understanding", 2/E, Addison-Wesley, 1994
3. Foundations of Statistical Natural Language Processing by Christopher D. Manning, Hinrich Schutze, MIT Press.
4. Statistical Language Learning by Charniack, Eugene, MIT Press, 1993.
5. The Handbook of Computational Linguistics and Natural Language Processing, Alexander Clark, Chris Fox, Shalom Lappin.
6. Steven Bird, Natural Language Processing with Python, 1st Edition, O'Reilly, 2009.

Course Outcome (CO):

CO No	Course Outcome
CO1	Understand the fundamental concept of NLP, Regular Expression, Finite State Automata along with the concept and application of word tokenization, normalization, sentence segmentation, word extraction, spell checking in the contextof NLP.
CO2	Understand the concept of Morphology such as Inflectional and Derivational Morphology and different morphological parsing techniques and scope of ambiguityand it’s resolution.
CO3	Understand the concepts of pragmatics, lexical semantics, lexical dictionary such as WordNet, lexical computational semantics, distributional word similarity and concepts related to the field of Information Retrieval in the context of NLP.
CO4	Understand the concepts of Semantic Roles; Word Sense Disambiguation;Multilinguality; Metaphors; Coreferences. Discourse, Dialogue and Conversational agents, Natural Language Generation, Machine Translation.
CO5	Understand the concepts related to language modeling with introduction to N-grams, chain rule, smoothing, spelling and word prediction and their evaluation along with the concept of Markov chain, HMM, Forward and Viterbi algorithm, POS tagging.
CO6	Describe and apply concepts of discourse machine translation, summarization and question answering to solve problems in NLP.

CO-PO Mapping (Rate on a scale of 1 to 3):

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	1	1	1	1	1	1
CO2	2	2	2	2	1	2
CO3	2	2	3	2	1	3
CO4	2	3	3	3	3	3
CO5	3	3	3	3	3	3
CO6	3	3	3	3	3	3
Total	13	14	15	14	12	15
Average	2.16	2.3	2.5	2.3	2	2.5
Eq. Average Attainment	2	2	3	2	2	3

4. Soft Computing	
L T P 4 , 0 , 0 : **4 Credits**	Prerequisites: *None*

Course Objectives:

1. To develop students' skill in neuro-fuzzy engines to handle machine learning in presence of uncertainty.
2. To provide solutions to real world problems with approximate reasoning using fuzzy logic.
3. To instill the scope of optimization in engineering design using evolutionary computation.
4. To demonstrate the scope of the subject in all aspects of science, humanities, and engineering.
5. To emphasize on the necessity of soft techniques in engineering industry, where mathematically hard techniques are difficult to realize in absence of sufficient data.

MODULE I

Introduction to Fuzzy sets, Fuzzy t- and s- norms, projection, cylindrical extension, Fuzzy relations, Implication relations, Fuzzy relational equations, Possibilistic reasoning, Fuzzy pattern recognition, Introduction to Fuzzy control and Fuzzy databases.

MODULE II

Boltzmann machine and Mean field learning-Combinational optimization problems using recurrent Neural network. Competitive Learning, Self-organizing maps, Growing cell structure, Principal component analysis.

MODULE III

Genetic Algorithm: Binary and real codes, Genetic programming, Particle swarm optimization, Differential Evolution, Bacterial Foraging

MODULE IV

Hybridization of neuro-fuzzy, neuro-GA, neuro-swarm, neuro-evolution algorithms. Applications in Pattern Recognition, Robotics, and Image Processing.

MODULE V

Belief Networks: Pearl's Model for Distributed Approach of Belief Propagation and Revision in a causal network, Concepts of D-separation, Bayesian Belief Networks, Dempster-Shafer theory for Orthogonal summation of Beliefs, Data Fusion techniques, Uncertainty management using Belief Networks.

MODULE VI

Visual Perception: Marr's 2- and 1/2-Dimensional Vision, 3-D Vision, Camera Model, Perspective Projection Geometry, Inverse Perspective Projection Geometry, 3D Reconstruction from 2D Images by Kalman Filter and other Prediction Algorithms.

MODULE VII

Advanced Models of Reasoning: Soundness and Completeness issues of Resolution based Proof procedures in propositional and predicate logic, Herbrand's theorem and Lifting Lemma, Herbrand interpretation, Temporal Logic, Reasoning with Space and Time, Distributed Models of Reasoning using Petri Nets, and other graph theoretic approaches.

Text Books:

1. A. Konar, Computational Intelligence: Principles, Techniques, and Applications, Springer 2005
2. A. P. Engelbrecht, Computational Intelligence

References:

1. A. Konar, Artificial Intelligence and Soft Computing: Behavioral and Cognitive Modeling of the Human Brain, CRC Press, 2018.
2. A. K. Sadhu and A. Konar, Multi-Agent Coordination: A Reinforcement Learning Approach, Wiley- IEEE Press, 2021.
3. D. E. Goldberg, Genetic Algorithms in Search Optimization and Machine Learning, Addison Wesley,3rd edition.
4. S. Haykin, Neural Networks: A comprehensive foundation, Pearson, 1999.

CO-PO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	3	2	3	2	2
CO2	2	2	3	2	1	2
CO3	3	3	2	2	2	2
CO4	2	2	2	2	2	3
CO5	3	3	3	3	3	3
Total	13	13	12	12	10	12
Average	2.6	2.6	2.4	2.4	2	2.4
EquivalentAverage Attainment	3	3	2	2	2	2

5. REINFORCEMENT LEARNING	
L T P 3 - 1 - 0: 4 Credits	Prerequisites: *None*

Course Objectives:

1. Learn how to define RL tasks and the core principals behind the RL, including policies, value functions, deriving Bellman equations.
2. Understand and work with tabular methods to solve classical control problems.
3. Understand and work with approximate solutions (deep Q network based algorithms).
4. Learn the policy gradient methods from vanilla to more complex cases.
5. Explore imitation learning tasks and solutions.

Detailed Syllabus:

MODULE I

Foundation: Introduction and Basics of RL, Defining RL Framework and Markov Decision Process, Polices, Value Functions and Bellman Equations, Exploration vs. Exploitation, Code Standards and Libraries used in RL (Python/Keras/Tensorflow)

MODULE II

Tabular methods and Q-networks: Planning through the use of Dynamic Programming and Monte Carlo, Temporal-Difference learning methods (TD(0), SARSA, Q-Learning), Deep Q-networks (DQN, DDQN, Duelling DQN, Prioritised Experience Replay)

MODULE III

Policy optimization: Introduction to policy-based methods 10. Vanilla Policy Gradient 11. REINFORCE algorithm and stochastic policy search 12. Actor-critic methods (A2C, A3C) 13. Advanced policy gradient (PPO, TRPO, DDPG)

MODULE IV

Recent Advances and Applications: Model based RL, Meta-learning, Multi-Agent Reinforcement Learning, Partially Observable Markov Decision Process, Ethics in RL, Applying RL for real-world problems

Text Books:

1. Sutton, Richard S., and Andrew G. Barto. “Reinforcement learning: An introduction,” First Edition, MIT press.
2. Sugiyama, Masashi. “Statistical reinforcement learning: modern machine learning approaches,” First Edition, CRC Press.

3. Boris Belousov, Hany Abdulsamad, Pascal Klink, Simone Parisi, and Jan Peters "Reinforcement Learning Algorithms: Analysis and Applications,"First Edition, Springer

Reference Books:

1. Lattimore, T. and C. Szepesvári. "Bandit algorithms," First Edition, Cambridge University Press.
2. Alexander Zai and Brandon Brown "Deep Reinforcement Learning in Action," First Edition, Manning Publications.
3. Li, Yuxi, "Deep Reinforcement Learning", https://arxiv.org/pdf/1810.06339.pdf

Course Outcomes (CO):

Course Outcome No	Course Outcome
CO1	Learn how to define RL tasks and the core principals behind the RL, including policies, value functions, deriving Bellman equations.
CO2	Understand and work with tabular methods to solve classical control problems.
CO3	Understand and work with approximate solutions (deep Q network-based algorithms)
CO4	Learn the policy gradient methods from vanilla to more complex cases.
CO5	Explore imitation learning tasks and solutions.

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	2	1	1	0	0
CO2	2	2	2	1	0	0
CO3	2	2	1	1	0	1
CO4	2	2	1	1	1	1
CO5	2	2	2	1	3	1
Total	10	10	7	5	4	3
Average	2	2	1.4	1	0.8	0.6
Eq. AverageAttainment	2	2	1	1	1	1

6. Intrusion Detection System	
L T P 4,0,0: 4Credits	Prerequisites: None

Course Objectives:

1. To introduce concepts in intrusion detection systems
2. To study and analysis, the different Intrusion Detection System Models.
3. To investigate the tools and methods of information assurance.
4. To investigate and simulate network and application security.
5. To explore the nature of secure intrusion detection system.

Detailed Syllabus:

MODULE-I

Introduction to IDS: Intruder types, intrusion methods, processes and detection, message integrity and authentication, honey pots

MODULE-II

IDS Models: General IDS model and taxonomy, data mining based IDS, Denning model, Framework for constructing features, and different models for intrusion detection systems, SVM, probabilistic, and statistical modelling, evaluation of IDS, cost sensitive IDS

MODULE-III

Network Security Threat Detection: NBAD, specification based and rate based DDOS, scans/probes, predicting attacks, network based anomaly detection, stealthy surveillance detection; defending against DOS attacks in scout, signature-based solutions, snort rules

MODULE-IV

Host based Threat Detection: Host-based anomaly detection, taxonomy of security flaws in software, self-modelling system calls for intrusion detection with dynamic window size.

MODULE-V

Secure Intrusion Detection Systems: Network security, secure intrusion detection environment, secure policy manager, secure IDS sensor, alarm management, intrusion detection system signatures, sensor configuration, signature and intrusion detection configuration, IP blocking configuration, intrusion detection system architecture.

Text Books:

1. J. Paul Guyer, "An Introduction to Intrusion Detection Systems," Create space Independent Publishers
2. Gerard Blokdyk, "Intrusion-detection System: How-to," Create space Independent Publishers.
3. Rash, M., Orebaugh, A. and Clark, G., "Intrusion Prevention and Active Response: Deploying Network and Host IPS," Syngress.
4. Endorf, C., Schultz E. and Mellander J., "Intrusion Detection and Prevention," McGraw-Hill.

Course Outcomes (CO):

CO Number	Course Outcome
CO1	Apply the intrusion detection system concepts for basic data science problem
CO2	Utilize the different Intrusion Detection System Models for data science network security and analysis.
CO3	Utilize the different open-source tools and methods information assurance for data science.
CO4	Demonstrate intrusion detection system using network security tool.
CO5	Implement Firewall design principles and identify various intrusion detection systems and be able to achieve highest system security

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	1	1	1	1	1	1
CO2	1	2	2	2	1	1
CO3	2	2	2	3	1	1
CO4	3	3	3	3	3	1
CO5	3	3	3	3	3	1
Total	10	11	11	12	9	5
Average	2	2.2	2.2	2.4	1.8	1
Eq. Avg. Attainment	2	2	2	2	2	1

7. COMPUTER VISION	
L T P 4, 0, 0: **4 Credits**	Prerequisites: None

Course Objectives:

1. To introduce students the fundamentals of image formation.
2. To introduce students the major ideas, methods and techniques of computer vision and pattern recognition.
3. To develop an appreciation for various issues in the design of computer vision and object recognition systems.
4. To provide the student with programming experience from implementing computer vision and object recognition applications.

Detailed Syllabus:

MODULE I

Digital Image Formation and low, level processing: Overview and State of the art, Fundamentals of Image Formation, Transformation: Orthogonal, Euclidean, Affine, Projective, etc, Fourier Transform, Convolution and Filtering, Image Enhancement, Restoration, Histogram Processing. Depth estimation and Multi camera views: Perspective, Binocular Stereopsis: Camera and Epipolar Geometry, Homography, Rectification, DLT, RANSAC, 3D reconstruction framework, Auto calibration.

MODULE II

Feature Extraction: Edges , Canny, LOG, DOG, Line detectors (Hough Transform), Corners , Harris and Hessian Affine, Orientation Histogram, SIFT, SURF, HOG, GLOH, Scale, Space Analysis, Image Pyramids and Gaussian derivative filters, Gabor Filters and DWT.

Image Segmentation: Region Growing, Edge Based approaches to segmentation, Graph, Cut, Mean, Shift, MRFs, Texture Segmentation, Object detection.

MODULE III

Motion Analysis: Background Subtraction and Modeling, Optical Flow, KLT, Spatio, Temporal Analysis, Dynamic Stereo, Motion parameter estimation.

Shape from X: Light at Surfaces, Phong Model, Reflectance Map, Albedo estimation, Photometric Stereo, Use of Surface Smoothness Constraint, and Shape from Texture, color, motion and edges.

MODULE IV

Miscellaneous: Applications: CBIR, CBVR, Activity Recognition, computational photography, Biometrics, stitching and document processing, Modern trends, super-resolution, GPU, Augmented Reality, cognitive models, fusion and SR&CS.

Text Books:

1. Szeliski, R., Computer Vision: Algorithms and Applications, Springer, Verlag London .
2. Forsyth, A., D. and Ponce, J., Computer Vision: A Modern Approach, Pearson Education.

References:

1. Hartley, R. and Zisserman, A., Multiple View Geometry in Computer Vision Cambridge University Press.
2. Fukunaga, K., Introduction to Statistical Pattern Recognition, Academic Press, Morgan Kaufmann.

Course Outcomes (CO):

CO Number	Course Outcome
CO1	Describe different image representation, their mathematical representation and different data structures used.
CO2	Classify different segmentation algorithm for given input.
CO3	Create a 3D object from given set of images.
CO4	Detect a moving object in video using the concept of motion analysis.
CO5	Recognize the object using the concept of computer vision

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	1	1	1	1	1	1
CO2	1	2	2	2	1	2
CO3	2	2	2	3	2	2
CO4	3	3	3	3	3	3
CO5	3	3	3	3	3	3
Total	10	11	11	12	10	11
Average	2	2.2	2.2	2.4	2	2.2
Eq. Avg. Attainment	2	2	2	2	2	2

8. Information Retrieval	
L T P 4 , 0 , 0 : **4 Credits**	Prerequisites: *None*

Course objective pattern e:

1. To understand fundamental concepts of Information retrieval systems.
2. To understand the knowledge of data structures and indexing methods in information retrieval Systems.
3. To learn the evaluation of different indexing techniques.
4. To learn and develop indexing systems for audio and visual documents.
5. To learn the concept of searching of webs.

Detailed Syllabus:

MODULE I
Basic Concepts of IR, Data Retrieval & Information Retrieval, IR system block diagram. Automatic Text Analysis, Luhn's ideas, Conflation Algorithm, Indexing and Index Term Weighing, Probabilistic Indexing, Automatic Classification, Measures of Association, Different Matching Coefficient, Classification Methods, Cluster Hypothesis. Clustering Algorithms, Single Pass Algorithm, Single Link Algorithm, Rochhio's Algorithm and Dendograms.

MODULE II
File Structures, Inverted file, Suffix trees & suffix arrays, Signature files, Ring Structure, IR Models, Basic concepts, Boolean Model, Vector Model, and Fuzzy Set Model. Search Strategies, Boolean search, serial search, and cluster based retrieval, Matching Function.

MODULE III
Performance Evaluation, Precision and recall, alternative measures reference collection (TREC Collection), Libraries & Bibliographical system, Online IR system, OPACs, Digital libraries , Architecture issues, document models, representation & access, Prototypes, projects & interfaces, standards.

MODULE IV
Taxonomy and Ontology: Creating domain specific ontology, Ontology life cycle Distributed and Parallel IR: Relationships between documents, Identify appropriate networked collections, multiple distributed collections, parallel IR, MIMD Architectures, Distributed IR, Collection Partitioning, Source Selection, and Query Processing.

MODULE V
Multimedia IR models & languages, data modelling, Techniques to represent audio and

visual document, query languages Indexing & searching, generic multimedia indexing approach, Query databases of multimedia documents, Display the results of multimedia searches, one dimensional time series, two-dimensional color images, automatic feature extraction.

MODULE VI
Searching the Web, Challenges, Characterizing the Web, Search Engines, Browsing, Mata searchers, Web crawlers, robot exclusion, Web data mining, Metacrawler, Collaborative filtering, Web agents (web shopping, bargain finder,..), Economic, ethical, legal and political Issues.

Text Books/References:

1. C.D. Manning, P. Raghavan, H. Schütze., Introduction to Information Retrieval. Cambridge UP, 2008.
2. R. Baeza-Yates, B. Ribeiro-Neto., Modern Information Retrieval. Addison-Wesley, 1999.
3. D.A. Grossman, O. Frieder., Information Retrieval: Algorithms and Heuristics., Springer, 2004.
4. I.H. Witten, A. Moffat, T.C. Bell., Managing Gigabytes., Morgan Kaufmann, 1999.
5. C.J. van Risjbergen., The Geometry of Information Retrieval., Cambridge UP, 2004.

Course Outcomes (CO):

CO Number	Course Outcome
CO1	Ability to understand the nature of information and retrieval requirements.
CO2	Ability to use knowledge of data structures and indexing methods in information retrieval systems.
CO3	Ability to evaluate performance of retrieval systems.
CO4	Ability to choose clustering and searching techniques.
CO5	Ability to crawl information and explain different types of search algorithms.

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	1	1	1	1	--	--
CO2	2	2	2	2	1	1
CO3	2	2	2	1	1	1
CO4	2	2	3	3	1	2
CO5	3	3	3	2	2	1
Total	11	11	11	9	5	5
Average	2.2	2.2	2.2	1.8	1	1
Eq. Avg. Attainment	2	2	2	2	1	1

9. Recommender Systems	
L T P 4 , 0 , 0 : 4 Credits	Prerequisites: *None*

Course Objective:

1. To learn the basic concepts for recommender systems.
2. To understand filtering algorithms and apply on recommendations.
3. To introduce different approaches of recommender systems.
4. To explore various types of recommender systems.

Detailed Syllabus:

MODULE I
Basic concepts for recommender systems, detailed taxonomy of recommender systems, Evaluation of recommender systems

MODULE II
Collaborative filtering algorithms: User-based nearest neighbour recommendation, Item-based nearest-neighbour recommendation, Model based and pre-processing based approaches, Attacks on collaborative recommender systems.

MODULE III
Content-based recommendation: High level architecture of content-based systems, Advantages and drawbacks of content based filtering, Item profiles, Discovering features of documents, Obtaining item features from tags, Representing item profiles, Methods for learning user profiles, Similarity based retrieval, Classification algorithms.

MODULE IV
Knowledge based recommendation: Knowledge representation and reasoning, Constraint based recommenders, Case based recommenders

MODULE V
Hybrid approaches: Opportunities for hybridization, Monolithic hybridization design: Feature combination, Feature augmentation, Parallelized hybridization design: Weighted, Switching, Mixed, Pipelined hybridization design: Cascade Meta-level, Limitations of hybridization strategies

MODULE VI
Evaluating Recommender System: General properties of evaluation research, Evaluation

designs, Evaluation on historical datasets, Error metrics, Decision-Support metrics, User-Centred metrics.

Reference Book:

1. Charu Aggarwal “Recommender Systems: The Textbook,” First Edition, Springer
2. Francesco Ricci, Lior Rokach, and Bracha Shapira “Recommender Systems Handbook,” First Edition, Springer
3. Rounak Banik “Hands-On Recommendation Systems with Python,” First Edition, Packt Publishing
4. Kim Falk “Practical Recommender Systems,” First Edition, Manning Publications
5. Deepak Agarwal and Bee-Chung Chen “Statistical Methods for Recommender Systems,” First Edition, Cambridge University Press

Course Outcomes (CO):

CO1	To learn the basic concepts for recommender systems.
CO2	To understand filtering algorithms and apply on recommendations.
CO3	To introduce different approaches of recommender systems.
CO4	To explore various types of recommender systems

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	1	1	--	--
CO2	2	2	2	2	--	--
CO3	2	3	3	2	2	2
CO4	2	2	3	2	2	2
Total	8	8	9	7	4	4
Average	2	2	2.25	1.75	1	1
Eq. Average Attainment	2	2	2	2	1	1

10. Deep Learning	
L T P 4 , 0 , 0 :4 Credits	Prerequisites: *None*

Course Objectives:

1. To introduce the idea of Artificial Neural Networks and their applications.
2. To study and implement different architectures of Artificial Neural Networks.
3. To study and implement various optimization techniques on Artificial Neural Networks.
4. To enable design and deployment of deep learning models for machine learning problems.

Detailed syllabus:

MODULE I
Introduction: Artificial Intelligence and Deep Learning-a historical perspective, Artificial neural networks, Shallow neural networks, Deep neural networks, gradient descent, forward and backpropagation, computational graphs, linear and non-linear activation functions.

MODULE II
Optimization techniques: Regularization, Dropout, Batch Normalization, Vanishing/Exploding gradients, Mini-batch gradient, Gradient descent with momentum, RMSprop, Adam optimization, Learning rate decay, Local optima, Global optima. Hyperparameter tuning,

MODULE III
Convolutional Neural Networks: Basic operations: padding, stride, pooling; Classic convolutional models: LeNet-5, AlexNet, VGG, Modern Deep Convolutional models: ResNet, GoogleNet; Inception Network, 1-D convolutions, Object detection and Face Recognition with CNN.

MODULE IV
Recurrent Neural Networks: Sequence modelling, Types of Recurrent Neural Networks, Backpropagation through time, Language modelling and sequence generation, Word Embeddings, vanishing gradients with RNNs, Long-Short Term Memory (LSTM), Gated Recurrent MODULEs (GRU), Bidirectional LSTMs, Sequence-to-Sequence model, Attention Mechanism, Transformer Network.

MODULE V
Advanced topics: Deep Reinforcement Learning, Generative Adversarial Networks, Generative vs. Discriminative models, Deep Convolution GANS, Autoencoders.

References:

1. Charu C. Aggarwal, Neural Networks and Deep Learning- A textbook, 2018, Springer.
2. Ian Goodfellow, Yoshua Bengio, Aaron Courville, ”Deep Learning (Adaptive Computation and Machine Learning series”, MIT Press.
3. Nikhil Buduma, Nicholas Locascio, “Fundamentals of Deep Learning: Designing Next Generation Machine Intelligence Algorithms”, O'Reilly Media.
4. Other online resources and research publications.

Course Outcomes (CO):

Course Outcome No.	Course Outcome
CO1	Students will be able to understand the mathematics and engineering sciences behind functioning of artificial neural networks.
CO2	Students will be able to analyze the given dataset and data attributesfor designing a neural network-based solution.
CO3	Students will be able to identify different neural network architectures, neural network optimization techniques, and apply them on different problem domains.
CO4	Students will be able to design and deploy deep learning solutions for real-world applications with popular deep learning tools.

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	2	2	1	1	--
CO2	2	2	1	2	2	--
CO3	3	3	3	3	2	1
CO4	3	3	2	2	2	2
Total	10	10	8	8	7	3
Average Attainment	2.5	2.5	2	2	1.75	0.75
Eq. Average Attainment	3	3	2	2	2	1

11. Data Visualization	
L T P 4 , 0 , 0 : 4 Credits	Prerequisites: *None*

Course Objectives:

1. This course is all about data visualization, the art and science of turning data into readable graphics.
2. How to design and create data visualizations based on data available and tasks to be achieved.
3. This process includes data modeling, data processing (such as aggregation and filtering), mapping data attributes to graphical attributes, and strategic visual encoding based on known properties of visual perception as well as the task(s) at hand.
4. Students will also learn to evaluate the effectiveness of visualization designs, and think critically about each design decision, such as choice of color and choice of visual encoding.

Detailed Syllabus:

MODULE I
Foundation: Importance of analytics and visualization in the era of data abundance, 2-D Graphics, 2-D Drawing, 3-D Graphics, Photorealism, Non-Photorealism, The Human Retina, Perceiving Two Dimensions, Perceiving Perspective.

MODULE II
Visualization of Numerical Data: Data Mapping, Charts, Glyphs, Parallel Coordinates, Stacked Graphs, Tufte's Design Rules, Using Colours.

MODULE III
Visualization of Non-Numerical Data: Graphs and Networks, Embedding Planar Graphs, Graph Visualization, Tree Maps, Principal Component Analysis, Multidimensional Scaling, Packing.

MODULE IV
Visualization Dashboard: Visualization Systems, Database Visualization, Visualization System Design.

REFERENCE BOOKS

1. T. Munzner, Visualization Analysis and Design, CRC Press, 2015.
2. Edward Tufte, The Visual Display of Quantitative Information (2nd edition), Graphics Press.
3. Colin Ware, Information Visualization: Perception for Design (2nd edition), Morgan Kaufmann.
4. Alberto Cairo, The Functional Art: An Introduction to Information Graphics and Visualization, New Riders, Pearson Education.
5. Nathan Yau, Data Points: Visualization That Means Something, Wiley.
6. Charles D. Hansen and Chris R. Johnson, Visualization Handbook, Academic Press.
7. Will Schroeder, Ken Martin, and Bill Lorensen, The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics, Kitware Inc. Publishers.

Course Outcome (CO):

Course Outcome No.	**Course Outcome**
CO1	Students will be able to explain design and create data visualizations.
CO2	Students will be able to conduct exploratory data analysis usingvisualization.
CO3	Students will be able to use knowledge of perception and cognitionto evaluate visualization design alternatives.
CO4	Students will be able to apply data transformations such as aggregation and filtering for visualization.
CO5	Students will be able to explain and identify opportunities for application of data visualization in various domains.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	--	1	1	1	--	--
CO2	1	2	2	2	--	--
CO3	1	3	3	3	2	2
CO4	3	3	3	3	2	3
CO5	3	3	3	3	2	2
Total	8	12	12	12	6	7
Average	1.6	2.4	2.4	2.4	1.2	1.4
Eq. Average Attainment	2	2	2	2	1	1

12. Data Science in Bioinformatics	
L T P 4 , 0 , 0 : 4 Credits	Prerequisites: *None*

Course Objectives:

1. To provide exposure to the Data Science within the context of its importance in biology.
2. To learn various methodologies and techniques in biology using Data Science.
3. To learn various tools for bioinformatics data analytics.
4. To learn deep learning approaches for bioinformatics applications.
5. To learn and apply various data science models in biology.

Detailed Syllabus:

MODULE I
Need for Data Science in Biology and Healthcare, Visualization tools for biological and bioinformatics datasets, data handling, transformations of data.

MODULE II
Data Science in genomics, from genetics to genomes, Alignment, and phylogenetic trees.

MODULE III
Structural bioinformatics, Proteomics, Protein structure prediction, integrative structural modeling, and structure-based drug design.

MODULE IV
AI algorithms, statistical tools, graph algorithms for bioinformatics data analytics.

MODULE V
Deep learning algorithms in perspective of bioinformatics applications, GANs for biological applications, Whole-cell modeling approaches.

Text Books:

1. Arthur M. Lesk, "Introduction to Bioinformatics", Oxford University Press) (Fifth Edition)
2. Jeil Grus, "Data Science from Scratch: First Principles with Python", O'Reilly Media Inc. (Second Edition,)
3. Vince Buffalo, "Bioinformatics Data skills", O'Reilly Media Inc.
4. Neil C. Jones and Pavel A. Pevzner, "An introduction to Bioinformatics Algorithms", The MIT Press.

Course Outcome (CO):

Course Outcome No.	Course Outcome
CO1	To understand the importance of Data Science in biology.
CO2	To acquire knowledge of different data science techniques in biology.
CO3	Learn and apply various tools for bioinformatics data analytics.
CO4	Learn and applying deep learning approaches for bioinformatics applications.
CO5	To acquire knowledge on various data science models in biology.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	2	1	1	--	--
CO2	2	2	2	1	--	--
CO3	1	2	2	2	1	1
CO4	2	2	2	2	2	1
CO5	2	2	2	1	2	1
Total	9	10	9	7	5	3
Average	1.8	2	1.8	1.4	1	0.6
Eq. Average Attainment	2	2	2	1	1	1

13. Data Science for Decision Making	
L T P 4 , 0 , 0 : 4 Credits	Prerequisites: *None*

Course Objective:

1. To learn the concept of data driven decision making.
2. To learn basic data analysis.
3. To learn various issues of design of data driven experiments.
4. To understand and application of decision-making tools.
5. To learn and apply statistical analysis on data.

Detailed Syllabus:

MODULE-I
Fundamentals of Analytics: Introduction to data-driven decision making; general introduction to data driven strategy and its importance; use of examples and mini-case studies to illustrate the role of statistical analysis in decision making.

MODULE-II
Basic Data Analysis: Various types of data that are commonly collected by firms; methods to be used and inferences/insights that can be obtained depending on the type of data that are available (stated versus revealed preference, level of aggregation, cross- sectional, time series, panel data and so forth); use of frequency distributions, mean comparisons, and cross tabulation; statistical inferences using chi-square; t-test and ANOVA.

MODULE-III
Experimental Design and Natural Experiments: Issues of design of experiments and internal and external validity; case studies in marketing; economics; and medicine etc.; A-B testing; and circumstances that provide us with “natural” experiments.

MODULE-IV
Decision making tools: Regression analysis and its applications; use of regression output in forecasting; promotional planning and optimal pricing; multivariate analysis (unsupervised learning) cluster analysis; factor analysis decision trees; elastic nets and random forests.

MODULE-V
Case Studies: To understand the problem at an intuitive level; use of simple data analysis and visualization to verify (or falsify) the intuition; use of appropriate statistical analysis to present your arguments.

Text Books:

1. F.S. Hillier and G.J. Liberman "Introduction to Operations Research" Tata McGraw Hill Education Private Limited.
2. Gregory S. Parnel, Terry A. Bresnick, Steven N. Tani, Eric R. Johnson "Handbook of Decision Analysis", Wiley.
3. Emily Moberg and Igor Linkov "Multi-Criteria Decision Analysis: Environmental Applications and Case Studies", CRC Press, Taylor and Francis group.
4. Adiel Teixeira de Almeida, Emel Aktas, Sarah Ben Amor, João Luis de Miranda "Advanced Studies in Multi-Criteria Decision Making", CRC Press.

Course Outcome (CO):

Course Outcome No.	Course Outcome
CO1	Understanding the concept of data driven decision making.
CO2	To acquire the knowledge of basic data analysis.
CO3	Able to run data driven experiments and design.
CO4	Able to apply decision making tools.
CO5	Understand and apply statistical analysis on data

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**
CO1	1	1	2	2	--	--
CO2	1	2	2	2	--	--
CO3	1	2	2	2	2	1
CO4	2	2	2	2	2	1
CO5	2	2	2	2	1	1
Total	7	9	10	10	5	3
Average	1.4	1.8	2	2	1	0.6
Eq. Average Attainment	1	2	2	2	1	1

14. Social Network Analysis	
L T P 4 , 0 , 0 : 4 Credits	Prerequisites: *None*

Course Objectives:

1. To introduce the basic notions used for social network analysis.
2. To learn different graph models in social network.
3. To learn network topologies and various network analysis features.
4. To learn different models in social network analysis.
5. To do experiments with network structure and equilibrium.

Detailed Syllabus:

MODULE-I
Social Network Analysis: Preliminaries and definitions, Erdos Number Project, Centrality measures, Balance and Homophily.

MODULE-II
Random graph models: Random graphs and alternative models, Models of network growth, Navigation in social Networks.

MODULE-III
Network topology and diffusion, Contagion in Networks, Complex contagion, Percolation and information, Epidemics, and information cascades.

MODULE-IV
Cohesive subgroups, Multidimensional Scaling, Structural equivalence, Roles and positions, Ego networks, Weak ties, Structural holes.

MODULE-V
Small world experiments, small world models, Origins of small world, Heavy tails, Small Diameter, Clustering of connectivity

MODULE-VI
The Erdos-Renyi Model, Clustering Models, Preferential Attachment

MODULE-VII
Navigation in Networks Revisited, Important vertices and page rank algorithm, Towards rational dynamics in networks, Basics of game theory.

MODULE-VIII
Coloring and consensus, biased voting, network formation games, network structure and equilibrium, behavioral experiments, Spatial and agent-based models

Text Books:

1. Wasserman, Stanley, and Joseph Galaskiewicz. Advances in social network analysis: Research in the social and behavioral sciences. Sage Publications
2. Knoke, David, and Song Yang. Social network analysis. Sage Publications.
3. Carrington, Peter J., John Scott, and Stanley Wasserman, eds. Models and methods in social network analysis. Vol. 28. Cambridge university press.
4. Liu, Bing. "Social network analysis." In Web data mining, pp. 269-309. Springer, Berlin, Heidelberg.

Course Outcome (CO):

Course Outcome No.	Course Outcome
CO1	To introduce the basic notions used for social network analysis.
CO2	To learn different graph models in social network.
CO3	To learn network topologies and various network analysis features.
CO4	To learn different models in social network analysis.
CO5	To do experiments with network structure and equilibrium.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**
CO1	2	1	-	2	--	--
CO2	1	2	2	2	--	--
CO3	2	2	3	3	2	1
CO4	2	3	3	3	2	2
CO5	2	3	2	2	2	1
Total	9	11	10	12	6	4
Average	1.8	2.2	2	2.4	1.2	0.8
Eq. Average Attainment	2	2	2	2	1	1

15. Time Series Data Analysis	
L T P 3 , 1 , 0 :4 Credits	Prerequisites: *None*

Course Objectives:

1. To learn the concept and properties of time series data
2. To learn Autoregressive models and forecasting for time series data.
3. Analyzing time series data using R programming.
4. Learn various models of time series data.

Detailed syllabus:

MODULE-I
Basic Properties of time-series data: Distribution and moments, Stationarity, Autocorrelation, Heteroscedasticity, Normality.

MODULE-II
Autoregressive models and forecasting: AR, ARMA, ARIMA models.
Random walk model: non-stationarity and unit-root process, Drift and Trend models.

MODULE-III
Regression analysis with time-series data using R programming.
Principal Component Analysis (PCA) and Factor Analysis.

MODULE-IV
Conditional Heteroscedastic Models: ARCH, GARCH. T-GARCH, BEKK-GARCH.
Introduction to Non-linear and regime-switching models: Markov regime-switching models, Quantile regression, Contagion models

MODULE-V
Introduction to Vector Auto-regressive (VAR) models: Impulse Response Function (IRF), Error Correction Models, Co-integration. Introduction to Panel data models: Fixed-Effect and Random-Effect models.

Text Books:

1. Chris Brooks “Introductory Econometrics for Finance,” Fourth Edition, Cambridge University Press.
2. Ruey S. Tsay “Analysis of Time-series data,” Third Edition, Wiley
3. John Fox and Sanford Weisberg “An R Companion to Applied Regression,” Third Edition, SAGE
4. Yves Croissant and Giovanni Millo “Panel Data Econometrics with R,” First Edition, Wiley

Course Outcome (CO):

CO-No.	Course Outcome
CO1	To learn the concept and properties of time series data
CO2	To learn Autoregressive models and forecasting for time series data.
CO3	Analyzing time series data using R programming.
CO4	Learn various models of time series data.

CO-PO Mapping:

Course Outcome	PO-1	PO-2	PO-3	PO-4	PO-5	PO-6
CO-1	1	1	1	2	1	1
CO-2	2	2	1	2	1	1
CO-3	2	2	2	1	1	1
CO-4	2	3	2	2	2	1
Total	7	8	6	7	5	4
Average	1.75	2	1.5	1.75	1.25	1
Attainment	2	2	2	2	1	1

****