Kavayitri Bahinabai Chaudhari

North Maharashtra University, Jalgaon

School of Computer Sciences

Master of Science in Data Science

M. Sc. (Data Science)

Syllabus

[*under Academic Flexibility*]

Faculty of Science and Technology

With effect from July 2025-26

Master of Science in Data Science
[2-Year M. Sc. (Data Science) as Per NEP-2020]

PROGRAMME OBJECTIVES (POs):

1. **Broadly Educated and Versatile** - Able to draw upon foundational knowledge, learn, adapt, and successfully bring to bear analytical and computational approaches on changing societal and technological challenges.
2. **Inspiring and Collaborative** - Able to induce and contribute to diverse teams, expertise, and experiences.
3. **Innovative** - Drives scientific and societal advancement through technological innovation and entrepreneurship.
4. **Engaged** - Is and remains engaged with the academics, technical and scientific professional communities.

PROGRAMME SPECIFIC OUTCOMES (PSOs):

At the end of the program the postgraduate will be able to:

PSO No.	PSO	Cognitive level
MSCDS PSO.1	Apply Data Science Knowledge statistical method for Discovering patterns, use statistical techniques to make sense of the data.	3
MSCDS PSO.2	Extracting the data for predictions. Analyze a complex computing problem and apply principles of computing to identify solutions.	4
MSCDS PSO.3	Machine learning can help analyze human behavior and create products to improve customer experience. Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles.	5
MSCDS PSO.4	Acquire and apply new knowledge as needed, using appropriate learning strategies.	3

Program at a Glance

Name of the program (Degree)	:	M. Sc. (Data Science)
Faculty	:	Science and Technology
Duration of the Program	:	Two years (four semesters) full time.
Commencement of Year	:	A.Y. 2025-2026
Medium of Instruction and Examination	:	English
Nature of the Program	:	Self-Finance / Sufficiency
Eligibility	:	Mathematics up to 12th level and B.Sc. (Data Science) or B.Sc. (Computer Science) or B.Sc. (Information Technology) or BCA, B.E./B.Tech. in Computer Science and Engineering, or B. E. / B. Tech in Information Technology Any undergraduate with major or minor in data science.
Seat Reservation	:	Home University: 70%; Other Maharashtra University: 20%: Other State University: 10%
Entry requirements	:	A Bachelor's degree (Honours / Research) for the 1- year/2-semester Master's programme. • A Bachelor's degree for the 2-year / 4-semester Master's degree programme. • A 4-year Bachelor's Degree for the 1-year/2- semester master's programmme. •Admission will be based on the evaluation of documentary evidence. • A 1-year/2-semester Post-Graduate Diploma programme requires 3-year/6-semester bachelor's degree
Category Reservation	:	As per Maharashtra state government of rules. All the reservations shall be applicable to candidates belonging

		to Maharashtra state only subject to the fulfillment of the eligibility criteria specified by respective authorities from time to time.
Admission Criteria	:	As per the University graduation merit/Entrance Examination. A candidate is eligible for admission to the post graduate courses offered by School of Computer Science if he/she has secured at least 50% (or equivalent CGPA) marks at graduation (45% marks for backward class i.e. SC/ST/DT/NT/OBC/SBC category candidates from Maharashtra state).
Examination Pattern	:	60% (External Assessment) + 40% (Internal Assessment).
Passing Standard	:	40% Marks with separate head of passing for internal as well as external assessment.
Fee structure	:	Attached separately.
Evaluation mode	:	CGPA
Total Credits of the program	:	88 (58 core credits, 12 Elective core credit, 4 Credit for RM and 14 OJT/FP.

Master of Science in Data Science
[2-Year M. Sc. (Data Science) as Per NEP-2020]

Syllabus Structure as per NEP 2020

KAVAYITRI BAHINABAI CHAUDHARI NORTH MAHARASHTRA UNIVERSITY, JALGAON

School of Computer Sciences, M.Sc. (Data Science) PROGRAMME (A. Y. 2025-26)

Credit distribution structure for Two Years/One Year PG Degree and Ph. D. Programme

Year (2 Yr PG)	Level	Sem (2Yr)	Major (Core) Subjects		RM	OJT/FP	RP	Cumulative Credits	Degree
			Mandatory (MDS)	Elective (MDSE)					
I	6.0	Sem-I	MDS-101 (4) (T) MDS-102 (4) (T) MDS-103 (2) (T) MDS-104 (2) (P) MDS-105 (2) (P)	MDSE-106A (2) (T) MDSE-106B (2) (P) OR MDSE-107A (2) (T) MDSE-107B (2) (P)	MDS-416 (4)	-	-	22	PG Diploma (After 3-YrDegree)
		Sem-II	MDS-201 (4) (T) MDS-202 (4) (T) MDS-203 (4) (T) MDS-204(2) (T) MDS-205 (2) (P) MDS-206 (2) (P)	MDSE-207A (2) (T) MDSE-207B (2) (p) OR MDSE-208 (A) (2) (T) MDSE-208B (2) (p)	-	-	-	22	
Cum. Cr. For PG Diploma			32	8	4	-	-	44	
Exit option: PG Diploma (44 Credits) after Three Year UG Degree									
II	6.5	Sem-III	MDS-301 (4) (T) MDS-302 (4) (T) MDS-303 (4) (T) (NPTEL, Swayam) MDS-304 (2) (T) MDS-305 (2) (P) MDS-306 (2) (P)	MDSE-307A (2) (T) MDSE-307B (2) (P) OR MDSE-308A (2) (T) MDSE-308B (2) (P)	-	-	-	22	M.Sc. (DS) After 3-Yr UG Or PG Degree after4-Yr UG
		Sem-IV	MDS-401 (4) (T) (NPTEL, Swayam) MDS-402 (4) (T) (NPTEL, Swayam)	-	-	MDS-403A (14) (P) OR MDS-403B (14) (P)	-	22	
Cum. Cr. for 2 Yr PG Degree			26	4	-	14	-	44	
Total Cr. for 2 Yr PG Degree			58	12	4	14	-	88	
2 Years-4 Sem. PG Degree (88 credits) after Three Year UG Degree or 1 Year-2 Sem PG Degree (44 credits) after Four Year UG Degree									

Abbreviations: Yr.: Year; Sem.: Semester; OJT: On Job Training: Internship/ Apprenticeship; FP: Field projects; RM: Research Methodology; Research Project: RP; Cumulative Credits: Cum. Cr., T- Theory Course, P – Practical course, MDS- School Specific Core Course, MDSE- School Specific Elective Course

Note: The courses which do not have practical, 'P' will be treated as 'T'.

KAVAYITRI BAHINABAI CHAUDHARI NORTH MAHARASHTRA UNIVERSITY, JALGAON

School of Computer Sciences, PG DEGREE M. Sc. (Data Science) PROGRAMME (A. Y. 2025-26)

Credit distribution structure for Two Years/ One Year PG (M.Sc.) Degree Programme

Teaching and Examination Scheme, Master of Science (M. Sc)

M.Sc. (Level 6.0) Sem- I (Name of Courses for - Major, RM, OJT, RP courses)

Sr No	Course Category	Name of the course (Title of the Paper)		Total Credit	Hours/ Semester	Teaching Scheme (hrs/week)		Evaluation Scheme		
						Theory	Practical	Continuous Internal Evaluation (CIE) (CA)	End Semester Evaluation (ESE) (UA)	Duration of Examination (Hrs)
						T	P			
1	Major	MDS-101 (4) (T)	Mathematical Foundation for Data Science	4	60	4	-	40	60	3
		MDS-102 (4) (T)	Data Structures and Algorithms	4	60	4	-	40	60	3
		MDS-103 (2) (T)	Fundamental of Programming	2	30	2	-	20	30	2
		MDS-104 (2) (P)	Lab on Python for Data Science	2	60	-	4	20	30	2
		MDS-105(2) (P)	Lab on Data Structures and Algorithms	2	60	-	4	20	30	2
2	Elective (Any One Group)	MDSE-106 A (2) (T)	Java Programming	2	30	2	-	20	30	2
		MDSE-106 B (2) (P)	LAB on JAVA Programming	2	60	-	4	20	30	2
		MDSE-107 A (2) (T)	Full Stack Development	2	30	2	-	20	30	2
		MDSE-107 B (2) (P)	Lab on Full Stack Development	2	60	-	4	20	30	2
3	Research	MDS-108	Research Methodology	4	60	--	4	40	60	3
			Total	22	420	12	16	220	330	

KAVAYITRI BAHINABAI CHAUDHARI NORTH MAHARASHTRA UNIVERSITY, JALGAON

School of Computer Sciences, PG DEGREE M. Sc. (Data Science) PROGRAMME (A. Y. 2025-26)

Credit distribution structure for Two Years/ One Year PG (M.Sc.) Degree Programme

Teaching and Examination Scheme, Master of Science (M. Sc)

M.Sc. (Level 6.0) Sem- II (Name of Courses for - Major, RM, OJT, RP courses)

Sr No	Course Category	Name of the course (Title of the Paper)		Total Credit	Hours/ Semester	Teaching Scheme (hrs/week)		Evaluation Scheme		
						Theory	Practical	Continuous Internal Evaluation (CIE) (CA)	End Semester Evaluation (ESE) (UA)	Duration of Examination (Hrs)
						T	P			
1	Major	MDS-201 (4) (T)	Principals of Data Analytics and Visualization	4	60	4	-	40	60	3
		MDS-202 (4) (T)	Artificial Intelligence and Machine Learning	4	60	4	-	40	60	3
		MDS-203 (4) (T)	Database Management Systems	4	60	4	-	40	60	3
		MDS-204 (2) (T)	Software Engineering	2	30	-	4	20	30	2
		MDS-205 (2) (P)	Lab on Visual Analytics	2	60	-	4	20	30	2
		MDS-206 (2) (P)	Lab on Artificial Intelligence and Machine Learning	2	60	-	4	20	30	2
2	Elective	MDSE-207A (2) (T) MDSE-207B(2)(P) **OR** MDSE-208B (2) (T) MDSE-208B(2)(P)	Digital Image Processing Lab on Digital Image Processing OR Natural Language Processing Lab on Natural Language Processing	4	60	4	-	40	60	3
			Total	22	390	14	12	210	330	-

KAVAYITRI BAHINABAI CHAUDHARI NORTH MAHARASHTRA UNIVERSITY, JALGAON

School of Computer Sciences, PG DEGREE M. Sc. (Data Science) PROGRAMME (A. Y. 2025-26)

Credit distribution structure for Two Years/ One Year PG (M.Sc.) Degree Programme

Teaching and Examination scheme, Master of Science (M. Sc)

M.Sc. (Level 6.5) Sem- III (Name of Courses for - Major, RM, OJT, RP courses)

Sr No	Course Category	Name of the course (Title of the Paper)		Total Credit	Hours/ Semester	Teaching Scheme (hrs/week)		Evaluation Scheme		
						Theory	Practical	Continuous Internal Evaluation (CIE) (CA)	End Semester Evaluation (ESE) (UA)	Duration of Examination (Hrs)
						T	P			
1	Major	MDS-301 (4) (T)	Big Data	4	60	4	-	40	60	3
		MDS-302 (4) (T)	Advanced Machine Learning and Deep Learning	4	60	4	-	40	60	3
		MDS-303 (4) (T)	Swayam / NPTEL (Online Course)	4	60	4	-	40	60	3
		MDS-304 (2) (T)	Data Warehousing and Data Mining	2	30	2	-	20	30	2
		MDS-305 (2) (P)	Lab on Big Data	2	60	-	4	20	30	2
		MDS-306 (2) (P)	Lab on Advanced Machine Learning and Deep Learning	2	60	-	4	20	30	2
2	Elective	MDSE-307A (2) (T) MDSE-307B (2) (P) **OR** MDSE-308A (2) (T) MDSE-308B (2) (P)	Computer Vision LAB on Computer Vision OR Large Language Model LAB on Large Language Model	2 2	30 60	2 -	- 4	20 20	30 30	2 2
			Total	22	420	16	12	220	330	--

Note: Major Courses MDS-303 must be completed from NPTEL /Swayam Online. The Subject will be given before the commencement of the semester

KAVAYITRI BAHINABAI CHAUDHARI NORTH MAHARASHTRA UNIVERSITY, JALGAON

School of Computer Sciences, PG DEGREE M. Sc. (Data Science) PROGRAMME (A. Y. 2025-26)

Credit distribution structure for Two Years/ One Year PG (M.Sc.) Degree Programme

Teaching and Examination scheme, Master of Science (M. Sc)

M.Sc. (Level 6.5) Sem- IV (Name of Courses for - Major, RM, OJT, RP courses)

Sr No	Course Category	Name of the course (Title of the Paper)		Total Credit	Hours/ Semester	Teaching Scheme (hrs/week)		Evaluation Scheme		
						Theory	Practical	Continuous Internal Evaluation (CIE) (CA)	End Semester Evaluation (ESE) (UA)	Duration of Examination (Hrs)
						T	P			
1	Major (NPTEL/ Swayam)	MDS-401	Swayam / NPTEL (Online Course)	4	60	4	-	-	100	3
		MDS-402	Swayam / NPTEL (Online Course)	4	60	4	-	-	100	3
2	FP/OJT, RP	MDS-403A OR MDS-403B	Industrial Training OR Inhouse Research Project	14	420	-	30	100	200	3
			Total	22	420	-	30	100	400	--

Note: Major Courses MDS-401 and MDS-402 must be completed from NPTEL /Swayam Online. The Subject will be given before the commencement of the semester

SEMESTER I

*Course Code: MDS-***101**	**Mathematical Foundation for Data Science (4)(T)**	*ClockHours:***60** *Total Marks:***100**

Course Objectives:

1. To get familiar and understand the fundamental notions in discrete mathematics, statistics and probability theory in problem solving.
2. To understand and demonstrate the basic concept required for data science.
3. To identify the basic properties of decision trees and Markov chain.

Unit-I	**[20]**	Max Marks:**25**

Linear Algebra: Systems of Linear Equations, Matrices, Inverse and Transpose, Multiplication by a Scalar, Compact Representations of Systems of Linear Equations.

Analytic Geometry: Norms, Inner Products, Lengths and Distances, Angles and Orthogonality, Orthonormal Basis, Orthogonal Complement, Inner Product of Functions, Orthogonal Projections, Rotations.

Basics of vectors: Direction, Magnitude, Addition, Subtraction, Dot Product (a Scalar Quantity), Unit vector, Projection of a vector, Angle between two vectors, Orthogonal vectors, Linear Combination of vectors, Sigmoid function.

Unit-II	**[10]**	Max Marks:**20**

Statistics: Introduction to Statistics, Classification of Data, Frequency Distribution, Graphic Representation: Histogram, Frequency Polygon, Cumulative Frequency Curves, Measures of Central Tendency: Mean, Median, Mode, Measures of Dispersion: Range, Quartile Deviation, Standard Deviation.

Unit-III	**[15]**	Max Marks:**25**

Discrete Probability: An Introduction to Discrete Probability, Finite Probability, Classical definition of Probability, Events and their Outcomes, Rules of Probability, Probability axioms, Probabilities of Complements and Unions of Events, Random variables (discrete and continuous), Conditional Probability, independence and Bayes theorem, Distribution of a random vector, Probability mass function, Probability density function and distribution function. Distributions: Binomial, Poisson, Uniform, Exponential, Normal.

Unit-IV	**[08]**	Max Marks:**15**

Decision Theory: Introduction, Elements of Decision-making problem, Decision models, Decision making under risk, Decision making under uncertainty, Maximin or Minimax criterion, Maximax or Minimin criterion, Hurwicz Alpha criterion, Laplace criterion, Minimax Regret criterion

Unit-V	**[07]**	Max Marks:**15**

Markov Chains: An Introduction to Markov Chains, Transition Probabilities, Transition Matrix, Transition Diagram, Probability Tree Diagram, State Probabilities, Steady State Probabilities, discrete and continuous Markov models, Chapman-Kolmogorov equation.

References:

1. Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Mathematics for Machine Learning, , Cambridge University Press, 2020. ISBN-*10. 110845514X* · ISBN-13. 978-1108455145
2. Ranjeet Chitale (1 January 2018), Statistical and Quantitative Methods, , Nirali Prakashan; Fifteenth edition ASIN: B07LG72157, ISBN13: 9788190693585.

3. Kenneth H. Rosen, Discrete Mathematics and Its Applications Eighth Edition, McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121. Copyright c 2019 by McGraw-Hill Education. ISBN 978-1-259-67651-2

Course Outcome:

1. Ability to distinguish between the notion of discrete and continuous mathematical structures
2. Ability to construct and represent data using statistical measures.
3. Application of mathematical, statistical and probabilistic techniques towards problem solving

Course Code: MDS-**102**	**Data Structures and Algorithms (4) (T)**	*ClockHours:***60** *Total Marks:***100**

Course Objectives:

1. To impart the basic concepts of data structures and algorithms
2. To understand basic concepts about array, stacks, queues, linked lists, trees and graphs
3. To impart knowledge of advance topics like AVL Trees, B Trees, B* and B+ Trees
4. To understand concepts about searching and sorting techniques
5. Apply hashing concepts for a given problem
6. To understanding about writing algorithms and step by step approach in solving problems with the help of fundamental data structures

Unit-I	**[08]**	Max Marks:**10**

Introduction to Data Structures and Algorithms:
Algorithmic Notation: Format Conventions, Statement and Control Structures. Time and Space Analysis: Data types and Abstract data types, Types of Data structures; Primitive, Non primitive, Linear and Nonlinear Data structures

Unit-II	**[08]**	Max Marks:**20**

Array, Stack and Queues
Array: Storage representation, operations and applications (Polynomial addition and subtraction) **Stack:** operations and applications (infix, postfix and prefix expression handling), **Queue:** operations and applications, Circular Queues: operations and applications, Concept of Double ended Queue and Priority Queue, Linked representation of stack and queue.

Unit-III	**[10]**	Max Marks:**13**

Linked Lists: Operations and Applications of Linear linked list (Polynomial addition and subtraction), Circular linked list and Doubly linked list

Unit-IV	**[11]**	Max Marks:**21**

Trees: Binary Trees, Binary Tree: Representations, Operations (insert/delete), Traversal (inorder, preorder, postorder, level order), Threaded Binary Tree, Search Trees: AVL Tree, single and double rotations, B-Trees: insertion and deletion, Introduction to B+ and B* Trees

Unit-V	**[11]**	Max Marks:**18**

Graphs and Their Applications: Representation (Matrix/Adjacency) and Traversal (Depth First Search/Breadth First Search), Spanning Trees, Minimal Spanning Tree (Prim's and Kruskals's algorithm), Shortest Paths and All Pair Shortest Path, Dijkstra's, Floyd-Warshall Algorithms.

Unit-VI	**[12]**	Max Marks:**18**

Hashing, Searching & Sorting:

Hash Table: Hash Function, Collision and its Resolution, Separate Chaining, Open Addressing (linear probing, quadratic probing, double hashing), Rehashing, Extendible Hashing **Searching:** Linear Search and Binary Search (array/binary tree). **Sorting:** General Background, Sorting Techniques: Bubble Sort, Insertion Sort, Selection Sort, Quicksort, Merge sort, Heapsort and Radix Sort.

References:

1. Tremblay, J. & Sorenson, P.G., (2001), An Introduction to Data Structures with Application, Mcgraw Hill India, ISBN: 978-0074624715, 0074624717
2. Langsam, Y., Augenstein, M.J. & Tenenbaum A.M., (2015), Data Structures using C School of Computer Sciences, KBCNMU, Jalgaon M.Sc. (Computer Science) Syllabus-2019-20 Page 9 and C++, 2nd Edition, Pearson Education ISBN: 978-9332549319, 9332549311
3. Balagurusamy, E., (2013), Data Structures using C, 1st Edition, Mcgraw Hill Education, ISBN: 978-1259029547, 1259029549
4. Weiss, M.A., (2002), Data Structures and Algorithm Analysis in C, 2nd Edition, Pearson India, ISBN: 978-8177583588, 8177583581
5. Horowitz, E., Sartaj S. & Mehta, D. (2008), Fundamentals of Data Structures in C++, Universities Press ISBN: 978-8173716065, 8173716064 *
6. Lafore, R., (2003), Data Structures & Algorithms in Java, 2nd Edition, Pearson India, ISBN: 978-8131718124, 8131718123
7. Kruse, R., Tondo, C.L., Leung B., & Mogalla S, (2006), Data Structures and Program Design in C, Pearson India, ISBN: 978-8177584233.

Course Outcome:

1. Understand the concept of Dynamic memory management, data types, algorithms, Big O notation.
2. Understand data structures such as arrays, linked lists, stacks and queues, graphs, trees and hash tables.
3. Solve problem involving graphs, trees and apply different sorting and searching algorithms.Specify and identify deficiencies in existing protocols, and then go onto formulate new and better protocols;

*Course Code: M**DS***-**103**	**Fundamental of Programming (2) (T)**	*ClockHours:3***0** *Total Marks:***50**

Course Objectives:

4. To introduce the fundamental concepts of programming and problem-solving using Python.
5. To develop skills in writing efficient, readable, and modular Python programs.
6. To provide hands-on experience with Python libraries for data processing, file handling, and debugging.
7. To equip students with the ability to apply Python programming for real-world applications in various domains.

Unit-I	**[08]**	Max Marks:**15**

BASICS OF PYTHON SPYDER (TOOL): Introduction Spyder, Setting working Directory, Creating and saving a script file, File execution, clearing console, removing variables from environment, clearing environment, Commenting script files, Variable creation, Arithmetic and logical operators, Data types and associated operations.		
Unit-II	**[08]**	Max Marks:**10**
Sequence data types and associated operations: Strings, Lists, Arrays, Tuples, Dictionary, Sets, Range. **NumPy:** ndArray\		
Unit-III	**[08]**	Max Marks:**15**
Pandas dataframe and dataframe related operations on Toyota Corolla dataset: Reading files, Exploratory data analysis, Data preparation and preprocessing. **Data visualization on Toyoto Corolla dataset using matplotlib and seaborn libraries:** Scatter plot, Line plot, Bar plot, Histogram, Box plot, Pair plot. **Control structures using Toyota Corolla dataset:** if-else family, for loop, for loop with if break, while loop, Functions.		
Unit-IV	**[06]**	Max Marks:**10**
CASE STUDY: Regression 1. Predicting price of pre-owned cars Classification 1. Classifying personal income		
References Books: 1) Prof. Reghunathan Rengasamy, Python for Data Science, , IIT Madras. (Free Available Online on NPTEL) 2) Mark Lutz, "Learning Python", O'Reilly Media, 5th Edition, 2013, ISBN: 978-1449355739		
Course Outcome: 1. By the end of this course, students will be able to: 2. Understand the syntax, semantics, and constructs of the Python programming language. 3. Design and implement algorithms using Python to solve computational problems. 4. Develop Python programs incorporating modules, functions, file I/O, and error handling. 5. Demonstrate the ability to use Python libraries for data manipulation and application development.		

Course Code:MDS-**104**	**Lab on Python for Data Science (2) (P)**	*ClockHours:6***0** *Total Marks:***50**
Course Objectives: 1. To introduce Python programming techniques and libraries essential for data science applications. 2. To provide hands-on experience in data manipulation, analysis, and visualization using Python. 3. To develop proficiency in implementing machine learning models and statistical methods with Python.		

4. To enable students to solve real-world data-driven problems using Python-based tools and frameworks.

Assignments:

1. Installation of Spyder and Execution of simple Python program.
2. Create sequence data of strings, lists, tuples, dictionary, sets and range.
3. Perform some of the operations on sequence data like indexing, how to get the elements given an index etc.
4. Operations that can be performed on a sequence data: slicing, concatenation and multiplication.
5. Methods that we can apply on any sequential data: List and String Methods.
6. NumPy
   Write a NumPy program to test element-wise for NaN of a given array.
   Write a NumPy program to create an array of 10 zeros, 10 ones, and 10 fives.
   Write a NumPy program to create a vector with values ranging from 15 to 55 and print all values except the first and last.
   Write a NumPy program to compute the sum of all elements, the sum of each column and the sum of each row in a given array.
   Write a NumPy program to find missing data in a given array.
   Write a NumPy program to create a 4x4 array. Create an array from said array by swapping first and last, second and third columns.
   Write a NumPy program to multiply two given arrays of the same size element-by-element.
7. Reading and Writing Data:
   Read and Write into Spyder the commonly used file formats like dot csv format, the dot xlsx format and the .text format.
8. Pandas library:
   Import the data into Spyder. We will be looking at how to create a copy of original data. How to get the attributes of data; How to do indexing and selecting data.
9. Pandas Dataframes:
   Data types of variables in a dataframe, numeric data types and character data types, , how to check the data types of each column in your dataframe.
10. Control structures:
    if elif family, for loop and while loop. functions, how do we define a function using Python?
11. Working on using frequency tables two-way tables
12. Data Visualization:
    Create basic plots using the matplotlib library. The basic plots include scatter plot, histogram and bar plot.
    Create basic plots using seaborn library and the basic plots includes scatter plot, histogram, bar plot, box and whiskers plot and pair wise plots.
13. How to identify the missing values and some of the approaches to fill in those missing values

References Books:

1) Prof. Reghunathan Rengasamy, Python for Data Science, , IIT Madras. (Free Available Online on NPTEL)

2) Mark Lutz, Learning Python , O'Reilly Media, 5th Edition, 2013, ISBN: 978-1449355739

Course Outcome:

1. Apply Python libraries such as NumPy, Pandas, and Matplotlib for data manipulation and visualization.
2. Perform exploratory data analysis (EDA) to extract insights from datasets.
3. Implement machine learning algorithms using Python libraries like Scikit-learn and TensorFlow.
4. Develop Python-based solutions for data science problems, demonstrating practical applications of data preprocessing, modeling, and evaluation

*Course Code: M**DS***-**105**	**Lab on Data Structures and Algorithms (2) (P)**	*ClockHours:6***0** *Total Marks:***50**

Course Objectives:

1. Solve real-world problems by reasoning about data structure choices, choose appropriate implementations.
2. To make the students write various programs and ADTS for all data structures.
3. Students will learn to write, debug, and test large programs systematically

Assignments:

1. **Implementation of programs based on the following**
   - Arrays
   - Multidimensional Arrays, Matrices
   - Stacks, Polish Notation
   - Queues
   - Deques
   - Linear Linked List, Circular Linked List, Doubly Linked List
   - Polynomial Addition/Subtraction
2. **Implementation of programs based on Trees**
   - Binary Search Tree
   - In-order, Pre-order and Post-order Traversals
   - Heap Tree
   - Implementation of programs based on Graphs
   - Depth First Traversal
   - Breadth First Traversal
   - Obtaining Shortest Path (Dijkstra and Floyd-Warshall)
   - Minimum spanning tree (Kruskal and Prim)
3. **Implementation of programs for Hash Table, Searching and Sorting techniques**
   - Hash Table
   - Linear and Binary Search (using array)
   - Bubble sort
   - Selection sort
   - Insertion sort
   - Radix sort
   - Quick sort
   - Merge sort

1. • Heap sort

References Books:

As per MDS-102 Syllabus

Course Outcome:

As a result of completing this course, students will be able to:

1. Develop solutions for a range of problems using procedure oriented / object-oriented programming.
2. Choose the appropriate data structure and algorithm design method for a specified application.

Course Code: **MDSE-106A**	**Java Programming (2) (T)**	*Clock Hours:* **30** *Total Marks:* **50**

Course Objectives:

1) Understand Fundamental concepts of object-oriented programming using Java technology.
2) Java applications development using polymorphism, inheritance, interfaces and inner classes and multi-threading.
3) GUI applications and event driven applications development.

Unit-I	**[04]**	Max Marks:**5**

An Introduction to Java: History of Java, Features of Java (Java Buzz words), Obtaining Java Environment, Setting up Java Environment, Structure of the Java Program, Creating a Source File, Compiling the Source File into a .class file, Executing the Program, The Java Virtual Machine, Comments, Data types, variables, Keywords, Operators, Control Structures, Arrays

Unit-II	**[04]**	Max Marks:**5**

Introduction to OOPs: OOPs concepts, Predefined classes(String, StringBuffer), type casting, wrapper classes, Input and Output, User defined class, object creation and initialization, finalize() method, static fields and methods, this keyword, Access specifier Inner class, Packages

Unit-III	**[05]**	Max Marks:**10**

Inheritance, Polymorphism and interfaces: Dynamic Polymorphism (Method Overloading and Method Overriding), Static Polymorphism, final keyword, Superclass, Subclass, super keyword, Abstract classes, Methods with a Variable Number of Parameters, Enumeration Classes, Interfaces, Reflection

Unit-IV	**[05]**	Max Marks:**10**

Multithreading and Exceptions: Creating Thread, Multi-Tasking using Threads, Thread Synchronization or Thread Safe, Thread Class Methods, Thread Communication, Thread Properties, ThreadGroup, Thread States (Life-Cycle of a Thread), Exception handling (try, catch, finally), throws clause, throw clause, Types of Exceptions(built-in, user defined), Assertions

Unit-V	**[08]**	Max Marks:**12**

Graphics Programming and event handling : Introduction to swing and awt, Creating a Frame, Positioning a Frame, Displaying Information in a Component, Working with 2D Shapes, Color, Special Fonts for Text, JComponent class Methods, Creating Components in Swing (PushButton, Label, JComboBox Class , JList Class, JMenu Class), Layout Manager (Flow Layout, Border Layout, Card Layout, Grid Layout, GridBag Layout), Basics of Event Handling, Listeners and Listener Methods, Mouse Events, Keyboard Events, AWT Event Hierarchy

Unit-VI	**[04]**	Max Marks:**08**

Streams, Files and JDBC: Input and output stream, Reading and Writing Binary Data, Reading and Writing text Data, File Management (File Class), The Design of JDBC, JDBC Configuration, Executing SQL Statements, Query Execution Scrollable and Updatable Result Sets, Row Sets, Metadata, Transactions.

References:

1) Horstman Cay, Cornell Gary, Core JavaTM2, Vol.1&2, Seventh Edition, Pearson education. *ISBN-10. 0131118269* · ISBN-13. 978-0131118263
2) Herbert Schildt, The Complete Reference, Seventh Edition, Tata McGraw-Hill. *ISBN*: 978-0-07-226385-5
3) Steven Holzner, JAVA 2 Programming Black Book, Wiley India. ISBN-10. 1588800970 · ISBN-13. 978-1588800978
4) Ivor Horton, Beginning Java 2, JDK 5 Ed, Wiley India. ISBN · *9780764568749*, 9780764568749.

Course Outcome:

After completion of this course students shall be able to -

1. Create Java application development using polymorphism, inheritance, and inner classes.
2. Develop GUI interface and event driven applications.
3. Manipulate databases through java application.

Course Code: **MDSE-106B**	**LAB on Java Programming (2) (P)**	*Clock Hours:* **60** *Total Marks:* **50**

Course Objectives:

1) Solve real-world problems by reasoning about Java programming choices, choose appropriate implementations.
2) To make the students write various programs in java technology.
3) Students will learn to write, debug, and test large programs systematically.

Implementation of programs based on the following

- Write a program that demonstrates program structure of java with use of arithmetical and logical implementation.
- Write a program that demonstrates string operations using String and StringBuffer class.
- Write a program to demonstrate inner class and static fields.
- Write a program that demonstrates inheritance, polymorphism.
- Write a program that demonstrates 2D shapes on frames.
- Write a program that demonstrates color and fonts.
- Write a program to illustrate the use of various swing components.
- Write a program that demonstrates use of dialog box and menus.
- Write a program that demonstrates event handling for various types of events.
- Write a program to illustrate multithreading.
- Write a program to illustrate exception handling.
- Write a program to demonstrate the use of File class.
- Write a program that demonstrates JDBC on application.
- Write a program that demonstrate package creation and use in program.

Course Outcome:

After completion of this course students shall be able to-

1. Develop solutions for a range of problems using object-oriented programming.
2. Choose the appropriate data structure and algorithm design method for a specified application.
3. Apply practical knowledge on the applications using Java.

Course Code: **MDSE-107A**	**Full Stack Development (2) (T)**	*ClockHours:3***0** *Total Marks:***50**

Course Objectives:

1. Understand the fundamentals of web development
2. Develop responsive web applications
3. Implement dynamic web interactions
4. Understand Django framework and interact with MySQL databases for robust server-side functionality.
5. Apply full stack skills i.e., Integrate front-end and back-end technologies to develop real-world projects.

Unit-I	**[05]**	Max Marks:**10**

Introduction to HTML & CSS : HTML Basics: Overview of HTML structure and syntax. Basic HTML tags and attributes. HTML5 Features, Semantic and Non-Semantic Tags, New Input Types and Form Attributes, SVG, Canvas, Audio, and Video Tags, CSS attributes (ID, Class, Style, Title), CSS types (Inline, Internal, External), Basic CSS styling concepts, CSS Box Model: Introduction. Display Properties, Essential CSS selectors (ID, Class, Tag, Basic CSS combinators (Descendant, Child). Layout Techniques: CSS Flexbox, CSS Grid, CSS Pseudo-classes and Pseudo-elements: Introduction. CSS3 Features: Backgrounds, gradients, transitions, transformations.

Unit-II	**[10]**	Max Marks:**15**

Basics of Bootstrap, Javascript and React :

Introduction to Bootstrap: Overview of Bootstrap: Mobile-first framework. Responsive Design Principles: Basic understanding. Bootstrap Grid System: Introduction to containers, rows, and columns. Bootstrap UI Components, Typography, Tables, Images, Buttons: Introduction to Bootstrap components. Advanced Bootstrap Techniques: Customizing Grid Tiers and Breakpoints: Introduction. Using CSS Grid with Bootstrap.

Introduction to JavaScript: Variables, Data Types, Operators: Basic JavaScript concepts. Conditional Statements and Loops. Introduction to JavaScript functions. DOM Basics: Introduction to the Document Object Model. Basic event handling. Introduction to jQuery, AJAX with jQuery: Introduction to asynchronous requests.

Introduction to React: React Essentials: Basic components and JSX. State and Props: Introduction to component state and props

Unit-III	**[10]**	Max Marks:**13**

Introduction to Django :

Django Overview: Introduction to the Django framework, Installation and Setup, creating a virtual environment and setting up a Django project, Basic project structure. Introduction to Django Template Language (DTL), Variables, Filters, and Tags, Introduction to template inheritance, Introduction to Django models, Basic understanding of Django Model Fields and Field Options. Introduction to views, Mapping URLs in Django, Path and Include in URL Mappings. **Basic CRUD Operations:** Creating, reading, updating, and deleting data. Admin Panel: Creating a superuser and registering models. **Django Forms and Validation:** Django Forms: Introduction to form handling. Model Forms: Creating forms from models. Validation: Basic form validation techniques.

Unit-IV	**[05]**	Max Marks:**12**

Django REST Framework and Database Essentials:

Django REST Framework Introduction: Introduction to REST APIs. REST API Interaction with React. Serializers and Viewsets, Basic routing and permission concepts. Authentication and Authorization: Introduction to authentication methods.

Database Essentials: Introduction to accessing MySQL databases with Python. Basic DML, DDL and DQL Operations, Interacting with Databases Using Python: Introduction to Python Libraries for DB Access: Overview of popular libraries.

References Books:

1) Jon Duckett, HTML and CSS: Design and Build Websites, Wiley, 2011, ISBN-10. *1118008189* ; ISBN-13. 978-1118008188
2) Elisabeth Robson and Eric Freeman, Head First HTML and CSS, O'Reilly Media, 2012, ISBN: 9780596159900
3) Thomas A. Powell, HTML & CSS: The Complete Reference, McGraw-Hill, 2017, *ISBN-10. 9780070701946* ; ISBN-13. 978-0070701946
4) David Flanagan, JavaScript: The Definitive Guide, Orielly, 2020, *ISBN-10. 0596805527* ; ISBN-13. 978-0596805524
5) Jon Duckett, JavaScript and JQuery: Interactive Front-End Web Development, Wiley, 2014, ISBN-*10. 9781118531648* ; ISBN-13. 978-1118531648
6) Laura Lemay, Mastering HTML, CSS & JavaScript Web Publishing - Covers HTMLS5, CSS3, jQuery, BPB Publications, 2023 ISBN: 9788183335157

7) Jake Spurlock, Responsive Web Development – Bootstrap, O'Reilly Media, Inc., 2013. *ISBN*-10, 1449343910. *ISBN*-13, 978-1449343910.
8) Cory Gackenheimer, Introduction to React, Apress, 2015. *ISBN-10. 9781484212462* · ISBN-13. 978-1484212462
9) Nigel George, Mastering Django, Packt Publishing, 2016, ISBN-13 : 9781787281141
10) SQL AND Python Programming by Byran Johnson, Tyler MacDonald, 2019 ISBN-10. 1951764269 · ISBN-13. 978-1951764265

Course Outcome:

1. By the end of this course, students will be able to:
2. Create responsive web pages using HTML, CSS, and Bootstrap.
3. Implement dynamic front-end interactions Using JavaScript, jQuery, and ReactJS to enhance user experience.
4. Utilize Django to create server-side applications and interact with MySQL databases.
5. Integrate full stack technologies to build comprehensive web applications.
6. Troubleshoot and optimize web applications for better performance and user experience.

Course Code: **MDSE-107B**	**Lab on Full Stack Development (2) (P)**	*ClockHours:6***0** *Total Marks:***50**

Course Objectives:

- To introduce web development fundamentals:
- To develop responsive web pages:
- To implement client-side scripting:
- To build interactive web applications:

Assignments:

1. Create a basic HTML page with a header, paragraph, and image.
2. Apply basic CSS styles (color, font-size, background-color) using inline, internal, and external stylesheets.
3. Use CSS selectors (ID, Class, Tag) to style different elements and demonstrate the CSS box model by styling an element with padding, border, and margin.
4. Use CSS Flexbox to create a responsive layout and build a simple responsive grid layout using CSS Grid.
5. Create a simple animation using CSS transitions and apply CSS gradients to a background.
6. Create a responsive layout using Bootstrap's grid system and use Bootstrap UI components like buttons, tables, and images.
7. Customize Bootstrap's grid tiers and breakpoints and use CSS Grid with Bootstrap for advanced layouts.
8. Write a JavaScript program to manipulate DOM elements (e.g., change text color on click) and use JavaScript functions to perform basic calculations.
9. Create a simple event listener for mouse clicks and use jQuery to make an AJAX request.
10. Create a simple React component using JSX and use state and props in a React component.
11. Build a simple counter application using React state and pass props from a parent to a child component.
12. Install Django, set up a new project, and use Django Template Language (DTL) to display variables and apply filters.

13. Define a simple Django model, create a corresponding database table, and create a view to display model data and map it to a URL.
14. Implement basic CRUD operations (create, read, update, delete) for a model and use Django's admin panel to manage model instances.
15. Create a simple REST API using Django REST Framework and use serializers and viewsets to handle API requests.

Text and Reference Books:

1. Jon Duckett, HTML and CSS: Design and Build Websites, Wiley, 2011, ISBN-10. 1118008189 ; ISBN-13. 978-1118008188
2. Elisabeth Robson and Eric Freeman, Head First HTML and CSS, O'Reilly Media, 2012, ISBN: 9780596159900
3. Thomas A. Powell, HTML & CSS: The Complete Reference, McGraw-Hill, 2017, ISBN-10. 9780070701946 ; ISBN-13. 978-0070701946
4. David Flanagan, JavaScript: The Definitive Guide, Orielly, 2020, ISBN-10. 0596805527 ; ISBN-13. 978-0596805524
5. Jon Duckett, JavaScript and JQuery: Interactive Front-End Web Development, Wiley, 2014, ISBN-10. 9781118531648 ; ISBN-13. 978-1118531648
6. Laura Lemay, Mastering HTML, CSS & JavaScript Web Publishing - Covers HTMLS5, CSS3, jQuery, BPB Publications, 2023 ISBN: 9788183335157
7. Jake Spurlock, Responsive Web Development – Bootstrap, O'Reilly Media, Inc., 2013. ISBN-10, 1449343910. ISBN-13, 978-1449343910.
8. Cory Gackenheimer, Introduction to React, Apress, 2015. ISBN-10. 9781484212462 · ISBN-13. 978-1484212462
9. Nigel George, Mastering Django, Packt Publishing, 2016, ISBN-13 : 9781787281141
10. SQL AND Python Programming by Byran Johnson, Tyler MacDonald, 2019 ISBN-10. 1951764269 · ISBN-13. 978-1951764265

Course Outcome:

As a result of completing this course, students will be able to:

1. Create Basic Web Pages
2. Develop Dynamic Web Content
3. Build Responsive and Interactive Interfaces
4. Implement RESTful API.
5. Design Full-Stack Web Applications

Course Code: **MDS-108**	**Research Methodology (4) (T)**	*Clock Hours:* **60** *Total Marks:* **100**

Course Objectives:

1)To study and understand the research issues & challenges, research goals, scientific methods

2)To study Sampling, External Validity, Levels of Measurement, Scaling and Qualitative Measures. Data Preparation, Descriptive Statistics and Correlation; and Inferential Statistics

3)Reviewing Literature and Research Papers; Writing Research Papers, Thesis, Reports and Project Proposals Plagiarism and Copyrights.

Unit-I	**[08]**	Max Marks:**10**

Research Foundations: Research, Research Goals and Quality Research, Types of Research,

Variables, Hypotheses and Data; Structure, Positivism and Post-Positivism; Scientific Methods, Reasoning and Arguments; Mathematical Methods of Proof and Research Fallacies.

Unit-II	**[08]**	Max Marks:**15**

CS Research Context: Nature of Computer Science, Scientific Methods in Computer science, Types of Research in CS, Research Methods in Computer Science, Research Paradigms in CS, Grand Challenges for CS Research.

Unit-III	**[10]**	Max Marks: **12**

Measurements: Sampling, External Validity, Internal Validity Levels of Measurement, Scaling and Qualitative Measures.

Research Design: Types of Designs, Experimental Design, Probabilistic Equivalence, Hybrid Experimental Designs and Quasi-Experimental Design.

Unit-IV	**[11]**	Max Marks: **25**

Sampling: Concepts of Statistical Population, Sample, Sampling Frame, Sampling Error, Sample Size, Non-Response. Characteristics of a good sample. Probability Sample – Simple Random Sample, Systematic Sample, Stratified Random Sample & Multi-stage sampling. Determining size of the sample – Practical considerations in sampling and sample size.

Unit-V	**[11]**	Max Marks:**20**

Data Analysis: Data Preparation – Univariate analysis (frequency tables, bar charts, pie charts, percentages), Bivariate analysis – Cross tabulations and Chi-square test including testing hypothesis of association.

Unit-VI	**[12]**	Max Marks: **18**

Research Skills: Reviewing Literature and Research Papers; Writing Research Papers, Thesis, Reports and Project Proposals; Formatting, Appendices, Citation Formats and Style; General Conventions, Issues, Plagiarism and Copyrights.

References:

1) Kumar, Research Methodology: a step-by-step guide for beginners, , Pearson Education. ISBN-13: 978-9351501336 ISBN-10: 9789351501336.
2) Kothari C.K. (2004) 2/e, Research Methodology – Methods and Techniques (New Age International, New Delhi). ISBN, 8122415229, 9788122415223
3) Dawson, C, Practical Research Methods, UBSPD Pvt. Ltd. ISBN-13. 978-8174764485

Course Outcome:

After completion of this course students shall be able to-

1. Understand the basic concepts of research and its methodologies, identify appropriate research topics, select and define appropriate research problems and parameters.
2. Prepare a project proposal.
3. Organize and conduct research in a more appropriate manner, writing research report and thesis.

SEMESTER II

Course Code: **MDS-201**	**Principals of Data Analytics and Visualization (4) (T)**	*ClockHours:**60*** *Total Marks:**100***

Course Objectives:

Objective

To introduce the general areas of data analytics, machine learning and visualization tools and techniques for better understanding, analyzing and presenting data and drawing inferences based on it.

Unit-I	**[10]**	Max Marks:**16**

Introduction:

Importance of analytics and visualization in the era of data abundance, Review of probability, statistics and random processes, Brief introduction to estimation theory, Introduction to machine learning, supervised and unsupervised learning, gradient descent, overfitting, regularization etc.

Unit-II	**[15]**	Max Marks:**25**

Clustering Techniques and Supervised Classification Methods:

K-means, Gaussian mixture models and expectation-maximization, agglomerative clustering, evaluation of clustering - Rand index, mutual information-based scores, Fowlkes-Mallows index etc. Supervised classification methods: K-nearest neighbor, Naive Bayes, logistic regression, decision tree, support vector machine.

Unit-III	**[12]**	Max Marks:**20**

Regression:

Regression: Linear models, ordinary least squares, ridge regression, LASSO, Gaussian Processes regression Sparse coding and dictionary learning, orthogonal matching pursuit

Unit-IV	**[08]**	Max Marks:**14**

Introduction to artificial neural networks: (ANNs), deep NNs, convolutional neural network (CNN), and other recent topics

Unit-V	**[15]**	Max Marks:**25**

Data visualization: Basic principles, categorical and continuous variables. Exploratory graphical analysis, Creating static graphs, animated visualizations - loops, GIFs and Videos, Data visualization in Python and R, examples from Bokeh, Altair, ggPlot, ggplot2, gganimate, ImageMagick etc., Introduction to Visualization Toolkit (VTK) for 3D computer graphics, image processing and visualization, Visualization for deep learning

Reference Books

1. Jerome H. Friedman, Robert Tibshirani, and Trevor Hastie, The Elements of Statistical Learning, Springer, 2001. *ISBN*-10. 0387848576 · *ISBN*-13. 978-0387848570
2. Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2006. *ISBN*-10: 0-387-31073-8. *ISBN*-13: 978-0387-31073-2
3. David G. Stork, Peter E. Hart, and Richard O. Duda, Pattern Classification (2nd edition), Wiley, 2000. ISBN: 978-0-471-05669-0.

4. Edward Tufte, The Visual Display of Quantitative Information (2nd edition), Graphics Press, 2001. *ISBN* · 9780961392147, 0961392142
5. Colin Ware, Information Visualization: Perception for Design (2nd edition), Morgan Kaufmann, 2004. *ISBN*:978-1-55860-819-1.
6. Alberto Cairo, The Functional Art: An Introduction to Information Graphics and Visualization, New Riders, Pearson Education, 2013. ISBN-10 9780321834737, ISBN-13 978-0321834737.
7. Nathan Yau, Data Points: Visualization That Means Something, Wiley, 2013. *ISBN*-10. 9788126550845 · *ISBN*-13. 978-8126550845
8. Charles D. Hansen and Chris R. Johnson, Visualization Handbook, Academic Press, 2004. *ISBN*-13. 978-0123875822
9. Will Schroeder, Ken Martin, and Bill Lorensen, The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics, Kitware Inc. Publishers, 2004. ISBN-10 : 193093419X, ISBN-13 : 978-1930934191

Course Outcome:

At the end of the course, the students should be able to

- Parse a real-world data analysis problem into one or more computational components learned in this course,
- Apply suitable machine learning and/or visualization techniques and analyze the results obtained to enable optimal decision making.
- This would also act as a first course in data science which would provide necessary pre-requisites and knowledge to explore more specialized and involved topics in machine learning, analytics, statistics etc.

Course Code: MDS-**202**	**Artificial Intelligence and Machine Learning (4) (T)**	*ClockHours:***60** *Total Marks:***100**

Course Objectives:

1. To introduce the fundamental concepts and techniques of Artificial Intelligence (AI) and Machine Learning (ML).
2. To develop problem-solving skills using AI/ML models for classification, regression, clustering, and decision-making.
3. To provide practical knowledge of implementing AI/ML algorithms using programming tools and frameworks.
4. To explore the ethical, societal, and technical implications of deploying AI/ML systems in real-world scenarios.

Unit-I	**[05]**	Max Marks:1**0**
Introduction: History, Can Machines think?, Turing Test, Winograd Schema Challenge, Language and Thought, Wheels & Gears		
Unit-II	**[05]**	Max Marks:10

Introduction: Philosophy, Mind, Reasoning, Computation, Dartmouth Conference, The Chess Saga, Epiphenomena		
Unit-III	**[05]**	Max Marks:**05**
State Space Search: Depth First Search, Breadth First Search, Depth First Iterative Deepening		
Unit-IV	**[10]**	Max Marks:10
Heuristic Search: Best First Search, Hill Climbing, Solution Space, TSP, Escaping Local Optima, Stochastic Local Search		
Unit-V	**[04]**	Max Marks:**10**
Introduction: Statistical Decision Theory - Regression, Classification, Bias Variance		
Unit-VI	**[06]**	Max Marks:**10**
Linear Regression, Multivariate Regression, Subset Selection, Shrinkage Methods, Principal Component Regression, Partial Least squares		
Unit-VII	**[04]**	Max Marks:**10**
Linear Classification, Logistic Regression, Linear Discriminant Analysis		
Unit-VIII	**[03]**	Max Marks:**05**
Perceptron, Support Vector Machines		
Unit-IX	**[08]**	Max Marks:**10**
Neural Networks - Introduction, Early Models, Perceptron Learning, Backpropagation, Initialization, Training & Validation, Parameter Estimation - MLE, MAP, Bayesian Estimation		
Unit-X	**[05]**	Max Marks:**10**
Decision Trees, Regression Trees, Stopping Criterion & Pruning loss functions, Categorical Attributes, Multiway Splits, Missing Values, Decision Trees - Instability Evaluation Measures		
Unit-XI	**[05]**	Max Marks:**10**
Bootstrapping & Cross Validation, Class Evaluation Measures, ROC curve, MDL, Ensemble Methods - Bagging, Committee Machines and Stacking, Boosting		

References:

1) Deepak Khemani (2013). A First Course in Artificial Intelligence, McGraw Hill Education (India). ISBN-10 9781259029981, ISBN-13 978-1259029981

2) Elaine Rich and Kevin Knight (1991). Artificial Intelligence, Tata McGraw Hill. *ISBN*-13: 978-0-07-008770-5. *ISBN*-10: 0-07-008770-9.

3) Stuart Russell and Peter Norvig (2009). Artificial Intelligence: A Modern Approach, 3rd Edition, Prentice Hall. ISBN-13. 978-0*136042594*

4) Tom Mitchell (1997). Machine Learning. First Edition, McGraw- Hill, ISBN 10: 0070428077 ISBN 13: 9780070428072

5) Ethem Alpaydin (2009). Introduction to Machine Learning, Edition 2, The MIT Press. ISBN 978-0-262-01243-0

Course Outcome:

1. Understand the theoretical foundations and principles of AI and ML, including supervised, unsupervised, and reinforcement learning.

2. Design, implement, and evaluate AI/ML models using modern tools such as Python, TensorFlow, and Scikit-learn.
3. Analyze and preprocess datasets to build efficient and accurate AI/ML solutions.
4. Demonstrate the ability to critically assess the ethical and societal impact of AI/ML systems and apply them responsibly.

Course Code: **MDS-203**	**Database Management Systems (DBMS) (4) (T)**	*ClockHours:***60** *Total Marks:1***00**

Course Objectives:

Objective

1. Introduction to the basic concepts of database management systems, learning to design databases using ER modelling, and decomposing databases on functional dependencies.
2. Understand Relational databases and principles and Learn basics of SQL to construct queries using SQL.
3. Understand and become familiar with the basic issues of transaction processing, concurrency control and recovery systems.
4. Describe and discuss selected advanced database topics such as distributed database and NoSQL Database.

Unit-I: Introduction	**[06]**	Max Marks:**10**

Database system-concept and applications, Purpose of Database system, Characteristics of DBMS, Database Users, 1-tier, 2-tier, 3-tier, Distributed, Parallel and Cloud based architectures along with its advantages and disadvantages, Levels of Database Architecture, Data Models, Data-schemas and instances, Data Independence, Role and responsibilities of DBA.

Unit-II: Relational Databases and Query Processing	**[15]**	Max Marks:**30**

Basics of E-R Diagram, Structure of Relational Databases, Informal design guidelines, Database Schemas, Keys, Schema diagrams, Relational Query Languages, Relational Operation. Overview of SQL, SQL Data Definition, Basic Structure of SQL Queries, Basic Operations, Set Operations, Null Values, Aggregate Functions, Nested Sub queries, Modification of Databases. Join Expressions, Views, Transactions, Integrity Constraints, SQL data types and Schemas, Authorization, Overview of Dynamic SQL and SQL CLI. Functions and Procedures, Triggers. Relational Algebra fundamental and extended Operations. Tuple and Domain Relational Calculus.

Overview of Query Processing, Measuring Query Cost, Selection Operation, Sorting, Join Operation, Other Operations and Evaluation of Expression. Overview of Query Optimization, Transformation of Relational Expression, Choice of Evaluation Plan.

Unit-III: Transaction Management and Concurrency Control	**[12]**	Max Marks:**20**

Transaction Concept, Transaction States, Transaction Atomicity and Durability, Isolation, Levels of Isolation, Serializability.

Lock based Protocol, Timestamp based Protocol, Validation based Protocol, Deadlock Handling.

Unit-IV: Recovery System and Data Cleaning and Transformation	**[12]**	Max Marks:**20**

Failure Classification, Storage, Recovery and Atomicity, Recovery Algorithms, Buffer Management, Early lock release and logical undo operations.

Data Cleaning, Data Transformation, Data Integrity, Data Quality, Database Optimization and Integration with BI tools, Ethics and Data Privacy.

Unit-V: Advanced Topics in Databases:	**[15]**	Max Marks:**20**

Distributed Database: Overview, Homogeneous and Heterogeneous Databases, Sharding and Replication, Distributed Data Storage, Distributed Transactions, Commit Protocols, Concurrency Control, Cloud based Databases.

NoSQL Database: Introduction to NoSQL, Key Features, Differences between RDBMS and NoSQL, Types of NoSQL databases, NoSQL use cases. Comparison between NoSQL and SQL, Replication and Sharding, Distribution Models, Consistency in Distributed Data, CAP Theorem, Concept of ACID Vs BASE, Dealing with Transaction Consistency and Eventual Consistency

Reference Books

10. Michael Kifer, Arthur Bernstein, P.M, Lewis and P.K. Panigrahi (2011), "Database Systems: An Application Oriented Approach", Second Edition, Pearson Education, 2011, ISBN: 9788131703748.
11. C. J. Date, A. Kannan and S. Swaminathan (2006), "An Introduction to Database Systems", Eighth Edition, Pearson Education, 2006, ISBN:978-81-7758-556-8
12. Silberschatz, H.F. Korth, and Sudarshan (2011), "Database System Concepts", TMH Publications, Sixth Edition, 2011, ISBN: 978-007-132522-6
13. Ramez Elmasri, Shamkant B. Navathe (2011), "Fundamentals of Database Systems" Seventh Edition, Pearson Education, 2011, ISBN: 978-0-13-397077-7.
14. Sadalage, P. & Fowler, NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Wiley Publications,1st Edition ,2019, ISBN 978-0-321-82662-6

Course Outcome:

After completion of this course students shall be able to-

1. Describe the fundamental elements of relational database management systems
2. Apply the relational model, specify integrity constraints, and explain how to create a relational database using an ER diagram and normalization techniques.
3. Apply SQL to create, query and manipulate relational databases.
4. Apply and ensure properties of transaction to preserve database integrity.
5. Determine partitioning and distribution of data across networked nodes of a DBMS and data optimization in a distributed environment.
6. Understand importance of Data Cleaning, Transformation, Optimizations and will be aware of the BI tools available.
7. Understand the principles of Ethics and Data privacy.
8. Use NoSQL databases to store huge amounts of unstructured data, which is flexible, scalable, and distributed.

Course Code: **MDS-204**	**Software Engineering (2) (T)**	*ClockHours:3***0** *Total Marks:**50***

Course Objectives:

Objective

1. Explore the basic Software Engineering methods and their application, layered technology, and process framework of software process models.
2. Understanding software requirements, software feasibility study, Design models, functional and behavioral models.
3. Understanding of modularity, coding standards, verification and validation and software testing approaches.

Unit-I: Introduction and Process Model	**[08]**	Max Marks:**14**

Changing nature of Software, Software engineering, Software development life cycle (SDLC), Software Myths. Process Models: Generic process model, Prescriptive process models, Specialized process models.

Agile Process models: Adaptive software development, Scrum method, Crystal method, Feature driven development, Lean software development, Extreme programming (XP).

Unit-II: Requirement Analysis and Modelling	**[08]**	Max Marks:**14**

Requirement Engineering, Eliciting Requirements and developing use cases, Building requirement model, Types of requirements, Feasibility studies. Negotiating and Validating requirements. Requirement analysis, Scenario based modelling, UML models that supplements use case, Data modelling concepts, class-based modelling. Requirement modelling strategy, Flow oriented modelling, Creating Behavior model, Pattern for requirement modelling.

Unit-III: Design Concept	**[06]**	Max Marks:**10**

Design process, Design Concept: Abstraction, Architecture, Pattern, Separation of concept, Modularity, Information hiding, Functional independence, Refinement, Aspects, Refactoring. Design Model: Data design element, Architectural design element, Interface design element, Component level design element, Deployment level design element

Unit-IV: Software Testing	**[08]**	Max Marks:**12**

Strategic approach to software testing, Test strategies for conventional software, Validation Testing, System testing, Software testing fundamentals, Internal and external view of testing, White box testing, Basic path testing, Control structure testing, Black box testing, Model based testing, Testing for specialized Environment, Architectures and applications.

Reference Books

1. R. S. Pressman, "Software Engineering: A Practitioner's Approach", McGraw-Hill International Edition, Seventh Edition, ISBN:978-007-126782-3.
2. Pankaj Jalote, "Software Engineering: A Precise Approach", Wiley India Pvt. Limited ISBN: 978-81-265-2311-5.
3. K. K. Aggarwal and Yogesh Singh, "Software Engineering", Third Edition, New Age International Publishers, ISBN:978-81-224-2360-0.

Course Outcome:

After completion of this course students shall be able to-

1. Translate end-user requirements into system and software requirements.
2. Apply software engineering principles and techniques, to develop, maintain and evaluate large-scale software
3. Produce efficient, reliable, robust and cost-effective software solutions
4. Work as an effective member or leader of software engineering teams and manage time, processes and resources effectively by prioritizing competing demands to achieve personal and team goals.

Course Code: **MDS-205**	**Lab on Visual Analytics (2) (P)**	*ClockHours:6***0** *Total Marks:***50**

Course Objectives:

Understand and implement different data analytics and visualization techniques using Python/R.

Implement following program using Python / R:

1. Write a program for cleaning and handling missing values in a dataset
2. Write a program to perform descriptive statistics: mean, median, mode, variance, etc.
3. Write a program for creating line charts, bar plots, scatter plots, and histograms. Plotting multiple graphs in a single figure.
4. Write a program for hypothesis testing using t-tests, ANOVA, and Chi-square tests.
5. Write a program for regression analysis, fitting a linear model and making predictions.
6. Write a program for binary classification using ML algorithms.
7. Write a program for model evaluation using accuracy, precision, recall, and F1-score.
8. Write a program to use the K-means clustering algorithm.
9. Write a program for text preprocessing using tokenization, stop word removal, and stemming
10. Write a program to work with time-series data.

Course Outcome:

At the end of the course, the students should be able to

- Proficiency in using popular data analytics libraries in Python/R and understanding practical aspects.

*Course Code:M**DS***-**206**	**MDS-206: Lab on Artificial Intelligence and Machine Learning (2) (P)**	*ClockHours:6***0** *Total Marks:***50**

Course Objectives:

1. To provide hands-on experience in implementing AI and ML algorithms using programming tools and frameworks.

2. To enable students to preprocess and analyze datasets for building AI/ML models.
3. To develop the ability to experiment with various AI/ML techniques for solving real-world problems.
4. To cultivate skills in evaluating the performance and accuracy of AI/ML models using appropriate metrics

1. **Basic Machine Learning with Scikit-learn**
    a. Objective: Familiarize with basic ML algorithms and Scikit-learn library.
    b. Task: Implement and evaluate classification algorithms (e.g., Logistic Regression, KNN, Decision Trees) on datasets like Iris or Wine.
2. **Linear Regression and Model Evaluation**
    a. Objective: Implement linear regression for predictive tasks and evaluate the model's performance.
    b. Task: Use datasets such as Boston Housing to implement linear regression and evaluate using metrics like RMSE and R^2.
3. **Support Vector Machines (SVM)**
    a. Objective: Implement SVM for both classification and regression tasks.
    b. Task: Use SVM on datasets like Iris or Breast Cancer for classification and evaluate using precision, recall, and F1 score.
4. **Random Forest and Feature Importance**
    a. Objective: Implement ensemble methods like Random Forest for classification tasks.
    b. Task: Use the Titanic dataset to implement Random Forest and identify important features impacting survival prediction.
5. **Naive Bayes Classifier**
    a. Objective: Understand and implement the Naive Bayes algorithm for text classification.
    b. Task: Use the Spam email dataset to classify emails as spam or not spam using Naive Bayes.
6. **Decision Trees and Overfitting**
    a. Objective: Train decision tree models and handle overfitting issues.
    b. Task: Implement decision trees for classification tasks and use pruning techniques to reduce overfitting on a dataset like the Adult Income dataset.
7. **Principal Component Analysis (PCA)**
    a. Objective: Apply dimensionality reduction techniques for visualization and model performance improvement.
    b. Task: Reduce the dimensions of a high-dimensional dataset (e.g., MNIST) using PCA and visualize the results in 2D.
8. **Artificial Neural Networks (ANN)**
    a. Objective: Build and train simple neural networks for classification tasks.
    b. Task: Implement an ANN using Keras or TensorFlow on datasets like MNIST for digit classification.
9. **K-Nearest Neighbors (KNN) for Classification and Regression**
    a. Objective: Explore KNN for both classification and regression problems.
    b. Task: Use KNN to classify species in the Iris dataset and predict continuous variables in the Boston Housing dataset.
10. **Text Classification using Naive Bayes**

a. Objective: Apply machine learning techniques to text classification tasks.
b. Task: Classify news articles or movie reviews (positive/negative) using Naive Bayes and vectorization methods like TF-IDF.

11. Feature Engineering and Selection

a. Objective: Improve model accuracy by engineering and selecting meaningful features.
b. Task: Implement feature scaling, transformation, and selection techniques on a dataset (e.g., Titanic or Heart Disease).

12. Model Evaluation and Cross-Validation

a. Objective: Understand model validation and evaluation techniques.
b. Task: Apply k-fold cross-validation on a classification task (e.g., Cancer Prediction) and evaluate model robustness.

13. Neural Networks for Image Classification

a. Objective: Implement Convolutional Neural Networks (CNNs) for image classification tasks.
b. Task: Train a CNN using a dataset like CIFAR-10 for classifying images into categories

References Books:

1) "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", Wes McKinney, O'Reilly Media, 2nd Edition, 2017, ISBN: 978-1491957669
2) "Python Machine Learning", Sebastian Raschka and Vahid Mirjalili, Packt Publishing, 3rd Edition, 2019, ISBN: 978-1788621750

Course Outcome:

1. Implement core AI and ML algorithms, including classification, regression, clustering, and optimization techniques.
2. Use Python-based tools like Scikit-learn, TensorFlow, and Keras to design and train AI/ML models.
3. Apply data preprocessing, feature selection, and dimensionality reduction techniques to enhance model performance.
4. Evaluate AI/ML models using performance metrics such as accuracy, precision, recall, and F1 score and interpret the results effectively.

Course Code: **MDS-207A**	**MDSE-207A: Digital Image Processing (2) (T)**	*ClockHours:***30** *Total Marks:***50**

Course Objectives:

1. Introduce the fundamental concepts of digital image processing.
2. Understand and implement image processing techniques for preprocessing images.
3. Develop a foundation for advanced topics in image understanding and analysis.

Unit-I	**[06]**	Max Marks:**10**

Introduction to Digital Image Processing: Overview of Image Processing: Applications and Challenges, Digital Image Fundamentals: Image Representation and Formats, Sampling and Quantization, Pixel Relationships and Connectivity

Unit-II	**[06]**	Max Marks:**10**

Image Enhancement in Spatial Domain : Gray Level Transformations: Contrast Stretching, Histogram Equalization, Basic Image Smoothing: Mean and Median Filters, Image Sharpening: Laplacian and High-Boost Filtering, Applications of Spatial Domain Techniques

Unit-III	**[06]**	Max Marks:**10**

Image Enhancement in Frequency Domain : Introduction to Fourier Transform for Images, Frequency Domain Filtering: Low-Pass and High-Pass Filters, Applications: Noise Reduction and Edge Enhancement

Unit-IV	**[04]**	Max Marks:**08**

Color Image Processing: Color models, Pseudo color Image processing, Color transformation and segmentation.

Unit-V	**[08]**	Max Marks:**12**

Image Segmentation and Morphological Processing: Edge Detection: Sobel, Prewitt, and Canny Operators, Thresholding Techniques: Global and Adaptive Thresholding, Morphological Operations: Erosion, Dilation, Opening, and Closing, Applications in Object Shape Analysis.

References Books:

1) Digital Image Processing, Fourth Edition by Rafel C. Gonzalez and Richard E. Woods, Pearson Education, ISBN-10, 9353062985. ISBN-13, 978-9353062989
2) Jain A.K., "Fundamentals of Digital Image Processing, Pearson education , *ISBN*-10. 0133361659 ; *ISBN*-13. 978-0133361650
3) Computer Graphics Principles and Practice" by John F. Hughes, Andries van Dam, Morgan McGuire, et al. *ISBN*-10. 0321399528 · *ISBN*-13. 978-0321399526
4) Online Resources: OpenCV Documentation, MATLAB Image Processing Toolbox.

Course Outcome:

1. By the end of this course, students will be able to:
2. Understand basic concepts image processing, image storage and types of transformations that can be applied to images.
3. Apply Image Restoration & Enhancement techniques, color image processing.
4. Process images using morphological and segmentation techniques

*CourseCode : M**DS***-**207B**	**MDSE-207B: Lab on Digital Image Processing(2) (P)**	*ClockHours:6***0** *Total Marks:***50**

Course Objectives:

1. Introduce the implementation basics of digital image processing.
2. Understand and implement enhancing images.
3. Develop a foundation for color image understanding and analysis of images.

Assignments

1. Write a program to Read/Write, and Display Images.
2. Write a program for converting Color Images to Grayscale,B/W.
3. Write a program to compute basic Image Statistics (Mean, Variance, Histogram).

4. Write a program to implement Contrast Stretching, Histogram Equalization, Negative Transformation and Log Transformation.
5. Write a program to implement Smoothing Filters: Mean, Median, and Gaussian Filters
6. Write a program to implement Sharpening Filters: Laplacian, Unsharp Masking, and High-Boost Filtering
7. Write a program to implement Gradient-Based Edge Detectors: Sobel, Prewitt, Canny Edge Detector
8. Write a program to implement Binary Image Processing: Erosion and Dilation
9. Write a program to implement Opening and Closing Operations
10. Write a program to implement Boundary Extraction and Region Filling
11. Write a program to implement Fourier Transform of Images
12. Write a program to implement Low-Pass and High-Pass Filtering in the Frequency Domain
13. Write a program to demonstrate Noise Models: Salt-and-Pepper, Gaussian Noise
14. Write a program to demonstrate Noise Removal using Averaging Filters, Median Filters, and Wiener Filters
15. Write a program to demonstrate Global and Adaptive Thresholding, Edge-Based Segmentation and Region-Based Segmentation.

References Books:

1) Digital Image Processing, Fourth Edition by Rafel C. Gonzalez and Richard E. Woods, Pearson Education, ISBN-10, 9353062985. ISBN-13, 978-9353062989
2) Jain A.K., "Fundamentals of Digital Image Processing, Pearson education , ISBN-10. 0133361659 ; ISBN-13. 978-0133361650
3) Computer Graphics Principles and Practice" by John F. Hughes, Andries van Dam, Morgan McGuire, et al. ISBN-10. 0321399528 · ISBN-13. 978-0321399526
4) Online Resources: OpenCV Documentation, MATLAB Image Processing Toolbox.

Course Outcome:

1. Apply basic concepts image processing and use basic operations on images.
2. Use Image Restoration & Enhancement technique.
3. Implement morphological and segmentation techniques

Course Code: **MDS-208A**	**MDSE-208A: Natural Language Processing (2) (T)**	*ClockHours:***30** *Total Marks:***50**

Course Objectives:

1. To provide a comprehensive understanding of the foundational concepts and techniques in Natural Language Processing (NLP).
2. To explore linguistic structures, text processing, and representation techniques for textual data.
3. To equip students with knowledge of Speech Processing and NLP.
4. To enable the application of NLP techniques to real-world problems such as machine translation, sentiment analysis, and question answering.

Unit-I	**[04]**	Max Marks:**06**
Introduction: NLP applications: Speech to Text(STT), Text to Speech(TTS), Story Understanding, NL Generation, QA system, Machine Translation, Text Summarization, Text classification, Sentiment Analysis, Grammar/Spell Checkers etc., challenges/Open Problems, NLP abstraction levels, Natural Language (NL) Characteristics and NL computing approaches/techniques and steps, NL tasks: Segmentation, Chunking, tagging, NER, Parsing, Word Sense Disambiguation, NL Generation, Web 2.0 Applications : Sentiment Analysis; Text Entailment; Cross Lingual Information Retrieval (CLIR).		
Unit-II	**[06]**	Max Marks:**08**
POS tagging Word Classes and Part-of-Speech tagging (POS), survey of POS tagsets, Rule based approaches (ENGTOWL), Stochastic approaches(Probabilistic, N-gram and HMM), unknown word handling, evaluation metrics: Precision/Recall/F-measure, error analysis		
Unit-III	**[06]**	Max Marks:**10**
Parsing Algorithms, Evidence for Deeper Structure; Top Down Parsing Algorithms, Parsing Algorithms, Probabilistic parsing; sequence labeling, PCFG, Probabilistic parsing; PCFG: Training issues, Arguments and Adjuncts, Probabilistic parsing; inside-outside probabilities, Indian Language Parsing in Paninian Karaka Theory		
Unit-IV	**[06]**	Max Marks:**10**
Speech : Phonetics, HMM, Morphology, Graphical Models for Sequence Labelling in NLP, Graphical Models for Sequence Labelling in NLP (contd.), Phonetics. Forward Backward probability; Viterbi Algorithm, Phonology. Consonants(place and manner of articulation) and Vowels		
Unit-V	**[04]**	Max Marks:**08**
Sentiment Analysis and Opinions on the Web, Machine Translation and MT Tools - GIZA++ and Moses, Text Entailment		
Unit-VI	**[04]**	Max Marks:**08**
Theories and approaches for Semantic Analysis, Meaning Representation, word similarity, Lexical Semantics, word senses and relationships, WordNet (English and IndoWordnet), Word Sense Disambiguation: Lesk Algorithm Walker's algorithm, Coreferences Resolution: Anaphora, Cataphora, Hexaphora		
References Books:		
1. Allen, James, Natural Language Understanding, Second Edition, Benjamin/Cumming, 1995. 2. Charniack, Eugene, Statistical Language Learning, MIT Press, 1993. 3. Jurafsky, Dan and Martin, James, Speech and Language Processing, Second Edition, Prentice Hall, 2008. 4. Manning, Christopher and Heinrich, Schutze, Foundations of Statistical Natural Language Processing, MIT Press, 1999. 4. Radford, Andrew et. al., Linguistics, An Introduction, Cambridge University Press, 1999. 5. Journals : Computational Linguistics, Natural Language Engineering, Machine Learning, Machine Translation, Artificial Intelligence. 6. Indurkhya, N., & Damerau, F. J. (Eds.). (2010). Handbook of Natural Language, Processing, 2nd Edition. New York: CRC Press Taylor and Francis Group, Boca Raton, London, New York. ISBN-10: 1420085921, ISBN-13: 978-1420085921		

7. Martin, J. H., & Jurafsky, D.(2013), Speech and Language Processing, Pearson Education, India; 2 edition, ISBN-10: 9332518416, ISBN-13: 978-9332518414
8. Manning, Christopher and Heinrich, Schutze(1999), Foundations of Statistical Natural Language Processing", MIT Press, ISBN-10: 0262133601, ISBN-13: 978-0262133609.
9. Akshar Bharati, Chaitanya, V., Kulkarni, A., & Sangal, R. (July 1997). Machine translation in Stages (Vol. 10 no. 3). Mumbai: NCST, Mumbai.
10. Bharati, A., Chaitanya, V., & Sangal, R. (1995). Natural Language Processing: A Paninian Perspective, New Delhi: Prentice Hall of India, ISBN 10: 8120309219, ISBN 13: 9788120309210.
11 Steven Bird, Edward Loper (2016),Natural Language Processing With Python, Ed. 2, O'Reilly Media,ISBN 1491913428, 9781491913420

Auxiliary Resources:

a. Web Links

1. https://see.stanford.edu/Course/CS224N
2. https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html

b. Video Links

1. http://www.nptelvideos.in/2012/11/natural-language-processing.html

https://www.youtube.com/playlist?list=PL6397E4B26D00A269

Course Outcome:

1. By the end of this course, students will be able to:
2. Understand the linguistic foundations of NLP, including syntax, semantics, and pragmatics.
3. Implement text preprocessing, feature extraction, and representation techniques like tokenization, stemming, and word embeddings.
4. Apply NLP tasks for speech processing tasks.
5. Develop and evaluate NLP solutions for real-world applications, considering computational efficiency, scalability, and ethical implications.

CourseCode: MDS-**208B**	**MDSE-208B: Lab on Natural Language Processing (2) (P)**	*ClockHours:6***0** *Total Marks:***50**

Course Objectives:

1. To provide hands-on experience in applying foundational and advanced Natural Language Processing (NLP) techniques to real-world datasets.
2. To enable students to preprocess and analyze textual data effectively using modern NLP libraries and frameworks.
3. To develop skills in implementing speech processing tasks for NLP.
4. To foster the ability to evaluate NLP models and address ethical concerns, such as bias and fairness, in text processing applications.

1. Text Preprocessing

- Objective: Learn the basics of cleaning and preparing text data for analysis.
- Task: Implement tokenization, stop word removal, stemming, and lemmatization on a text dataset (e.g., news articles or movie reviews).

2. Word Frequency Analysis

- Objective: Perform frequency-based analysis of text.

- Task: Create a word cloud and calculate term frequency (TF) and term frequency-inverse document frequency (TF-IDF) for a given text corpus.

3. Language Modeling
- Objective: Explore statistical language models.
- Task: Build unigram, bigram, and trigram models for a text dataset and calculate the probabilities of word sequences.

4. Sentiment Analysis
- Objective: Perform sentiment classification using traditional machine learning methods.
- Task: Use a dataset like IMDB movie reviews or Twitter sentiment data to classify text as positive, negative, or neutral using Naive Bayes or Logistic Regression.

5. Named Entity Recognition (NER)
- Objective: Identify entities such as names, dates, and locations in text.
- Task: Implement an NER system using libraries like SpaCy or fine-tune a transformer model for custom NER tasks.

6. Part-of-Speech (POS) Tagging
- Objective: Perform POS tagging on textual data.
- Task: Use libraries like NLTK or SpaCy to tag parts of speech in a given text and analyze the results.

7. Building a Chatbot
- Objective: Develop a simple conversational agent.
- Task: Use rule-based or retrieval-based approaches to create a chatbot for basic question answering or FAQs.

8. Text Generation
- Objective: Generate coherent and meaningful text using language models.
- Task: Fine-tune a GPT-based model for text generation in specific domains like storytelling or programming.

9. Dependency Parsing
- Objective: Analyze syntactic dependencies in sentences.
- Task: Use SpaCy or Stanford NLP to extract and visualize dependency trees for complex sentences.

10. Building Custom Word Embeddings
- Objective: Train custom word embeddings for domain-specific text.
- Task: Use Word2Vec or FastText to create embeddings for a custom dataset like product reviews or legal documents.

11. Sentiment Analysis Using Transformer Models
- Objective: Apply transformer-based models for sentiment classification.
- Task: Fine-tune a BERT or DistilBERT model on a sentiment dataset and evaluate its performance.

References Books:

1) **"Python Cookbook",** David Beazley and Brian K. Jones, O'Reilly Media, 3rd Edition, 2013, **ISBN**: 978-1449340377

2) **"Python Natural Language Processing",** Jalaj Thanaki, Packt Publishing, 2nd Edition, 2020, **ISBN**: 978-1838982751

3) **"Natural Language Processing with Python" (the NLTK Book),** Steven Bird, Ewan Klein, Edward Loper, O'Reilly Media, 1st Edition, 2009, **ISBN**: 978-0596516499
4) **"Hands-On Natural Language Processing with Python",** Rajdeep Dua, Manpreet Singh Ghotra, Karan Jain, Packt Publishing, 1st Edition, 2020, **ISBN**: 978-1800200237
5) **"Deep Learning for Natural Language Processing",** Palash Goyal, Sumit Pandey, Karan Jain, Apress, 1st Edition, 2018, **ISBN**: 978-1484237240
6) **"Mastering Natural Language Processing with Python",** Deepti Chopra, Nisheeth Joshi, Iti Mathur, Packt Publishing, 1st Edition, 2019, **ISBN**: 978-1788621752

Course Outcome:

1. Apply text preprocessing techniques such as tokenization, stemming, lemmatization, and word embeddings to prepare textual data for analysis.
2. Implement traditional and deep learning-based NLP models for tasks like sentiment analysis, named entity recognition, and text summarization using libraries such as NLTK, SpaCy, and TensorFlow.
3. Fine-tune and evaluate transformer-based models (e.g., BERT, GPT) for specific NLP applications.
4. Analyze ethical considerations in NLP, including bias detection and mitigation, while building models that handle diverse and inclusive text datasets.

SEMESTER III

Course Code: **MDS-301**	**Big Data (4) (T)**	*ClockHours:***60** *Total Marks:***100**

Course Objectives:

4. To introduce the current scenarios of big data and provide various facets of big data.
5. To provides opportunity to be familiar with the technologies playing key role in it and equips them with necessary knowledge to use them for solving various big data problems in different domains.
6. To Understand and introduce the Google file System.
7. To understand the Map Reduce Framework.
8. To Understand NoSQL and Searching and Indexing Big data.
9. To Know and Understand the Hadoop Framework and HDFS.

Unit-I Introduction to Big Data	**[10]**	MaxMarks:**20**

Introduction to Big data, Conventional Data vs Big data, Big data architecture Big data platforms, Nature of data, Analytic processes and tools, 5 V's of Big data, Big data analytics life cycle, Role of Distributed System in Big Data, Evolution of Big data, Best Practices for Big data Analytics, Big data characteristics, Validating, The Promotion of the Value of Big Data, Big Data Use Cases, Characteristics of Big Data Applications, Perception and Quantification of Value, Understanding Big Data Storage, A General Overview of High-Performance Architecture, Role of Data Scientist, Current Trend in Big Data Analytics.

Unit-II: Google File System and Map-Reduce Framework	**[15]**	Max Marks:**20**

Google File System (GPS) Architecture, Features of Google File System, Optimization for large scale data, Basics of functional programming, fundamentals of functional programming, Real world problems modeling in functional style, Map reduce fundamentals, Data flow (Architecture), Real world problems, Scalability goal, Fault tolerance, Optimization and data locality, Parallel Efficiency of Map-Reduce.

Unit-III: NoSQL , Searching and Indexing of Big Data	**[10]**	Max Marks:**20**

Structured and Unstructured Data, Taxonomy of NoSQL Implementation, Discussion of basic architecture of Hbase, Cassandra and MongoDb, Full text Indexing and Searching, Indexing with Lucene, Distributed Searching with elastic search.

Unit-IV: Pig, Hive, HBase	**[10]**	Max Marks:**20**

Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators. Introduction to Hive, Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions. Introduction to Hbase, HBasics, Concepts, Clients, Hbase, Analyzing big data with twitter, Big data for E-Commerce Big data for blogs, Review of Basic Data Analytic Methods using R. Hbase Versus RDBMS.

Unit-V: Hadoop	**[15]**	Max Marks:**20**
History of Hadoop, Introduction to Hadoop Environment, Hadoop I/O, Query languages for Hadoop, Hadoop and Amazon Cloud, Hadoop Ecosystem, Core Components, HDFS-Architecture, Using HDFS, Files, HDFS Design, Blocks, Name nodes and Data nodes, Basic File system Operations, Hadoop Specific File Types, Anatomy of a file read, Anatomy of a file write. Data Processing with MapReduce: Execution Pipeline, Runtime Coordination and Task Management in MapReduce, Designing MapReduce implementations: Using MapReduce as a framework for parallel processing.		
References: 1. Jeffrey Dean, Sanjay Ghemawat MapReduce:Simplified Data Processing on Large Clusters 2. Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung The Google File System 3. Tom White " Hadoop: The Definitive Guide" Third Edit on, O'reily Media, 2012. 4. Michael Minelli, Michelle Chambers and AmbigaDhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013. 5. Boris Lublinsky, Kevin T. Smith, Alexey Yakubovich ,Professional Hadoop Solutions. 6. http://wiki.apache.org/hadoop/ 7. https://www.geeksforgeeks.org/mapreduce-programming-model-and-its-role-in-hadoop/ 8. https://www.geeksforgeeks.org/google-file-system/ 9. https://archive.nptel.ac.in/courses/106/104/106104189/ 10. https://waytoeasylearn.com/learn/mapreduce-data-flow/		
Course Outcome: 4. Learn and understand Big Data with current scenarios of big data and provide various facets of big data. 5. Solving various big data problems in different domains. 6. Learn and understand and implement the Google File System and Map-Reduce Framework. 7. Use MongoDB to execute NoSQL Commands. 8. Understand and Configure Hadoop and perform File Management Tasks.		

Course Code: **MDS-302**	**MDS-302: Advanced Machine Learning and Deep Learning (4) (T)**	*ClockHours:6***0** *Total Marks:10***0**
Course Objectives: 1. To deepen understanding of advanced machine learning techniques and explore the principles of deep learning architectures.		

2. To equip students with the skills to design, implement, and optimize complex machine learning and deep learning models.
3. To introduce cutting-edge concepts like transfer learning, generative models, and reinforcement learning.
4. To analyze and address real-world challenges using advanced ML/DL methods while considering computational efficiency and scalability.

Unit-I	**[05]**	Max Marks: **08**
Introduction to ML, Performance Measures, Bias-Variance Trade off, Linear Regression.		
Unit-II	**[04]**	Max Marks: **08**
Bayes Decision Theory, Normal Density and Discriminant Function, Bayes Decision Theory - Binary Features, Bayesian Belief Network		
Unit-III	**[05]**	Max Marks: **08**
Parametric and Non- Parametric Density Estimation Parametric and Non- Parametric Density Estimation – ML and Bayesian Estimation, Parzen Window and KNN		
Unit-IV	**[05]**	Max Marks: **08**
Logistic Regression, Support Vector Machine Perceptron Criteria, Discriminative models, Support Vector Machines (SVM)		
Unit-V	**[05]**	Max Marks: **08**
Random Forest, Hidden Markov Model Logistic Regression, Decision trees, Hidden Markov Model (HMM)		
Unit-VI	**[05]**	Max Marks: **08**
Ensemble methods Ensemble methods: Ensemble strategies, boosting and bagging, Random Forest		
Unit-VII	**[04]**	Max Marks: **08**
Dimensionality Problem Dimensionality Problem, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA)		
Unit- VIII	**[05]**	Max Marks: **08**
Gaussian mixture model Concept of mixture model, Gaussian mixture model, Expectation Maximization Algorithm, K- means clustering		
Unit-IX	**[05]**	Max Marks: **08**
Clustering Fuzzy K-means clustering, Hierarchical Agglomerative Clustering, Mean-shift clustering		
Unit-X	**[05]**	Max Marks:**08**
Neural Network Neural network: Perceptron, multilayer network, backpropagation, RBF Neural Network, Applications		
Unit-XI	**[06]**	Max Marks:**10**
Introduction to deep neural network, Introduction to Deep Learning, Convolutional Neural Networks (CNN), Vanishing and Exploding Gradients in Deep Neural Networks, LeNet - 5, AlexNet, VGGNet, GoogleNet, and ResNet.		
Unit-XII	**[06]**	Max Marks:**10**
Recent Trends in Deep Learning Generative Adversarial Networks (GAN), Auto Encoders and Relation to PCA, Recurrent Neural Networks, U-Net, Applications and Case studies		
References Books:		

1) Tom Mitchell (1997). Machine Learning. First Edition, McGraw- Hill, ISBN 10: 0070428077 ISBN 13: 9780070428072
2) Ethem Alpaydin (2009). Introduction to Machine Learning, Edition 2, The MIT Press. ISBN 978-0-262-01243-0

Course Outcome:

1. Comprehend advanced machine learning techniques, including ensemble learning, support vector machines, and Bayesian approaches.
2. Understand and apply deep learning concepts such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers.
3. Develop, train, and optimize complex neural network models using frameworks like TensorFlow and PyTorch.
4. Solve real-world problems using advanced ML/DL methods and critically evaluate their performance, scalability, and ethical implications.

Course Code: **MDS-303**	**MDS-303: Swayam Online (4) (T)**	*ClockHours:6***0** *Total Marks:10***0**
Major Courses MDS-303 must be completed from NPTEL /Swayam Online. The Subject will be given before the commencement of the semester.		

Course Code: **MDS-304**	**Data Warehousing and Data Mining (2) (T)**	*ClockHours:3***0** *Total Marks:***50**

Course Objectives:

Objective

1. To understand the principles of Data warehousing and Data Mining.
2. To be familiar with the Data warehouse architecture and its Implementation.
3. To know the Architecture of a Data Mining system.
4. To understand the various Data preprocessing Methods.
5. To perform classification and prediction of data.

Unit-I	**[04]**	Max Marks:**06**

Introduction to Data Warehousing: Lifecycle of data, Types of data, Data warehouse and data warehousing, Differences between operational database and data warehouse, A multidimensional data model, OLAP operation in multidimensional data model, Conceptual modeling of data warehouse, Architecture of data warehouse, Data warehouse implementation, Data marts, Components of data warehouse, Need for data warehousing, Trends in data warehousing.

Unit-II	**[02]**	Max Marks:**04**

Introduction to Data Mining: Motivation for data mining, Introduction to data mining system, Data mining functionalities, KDD, Data object and attribute types, Statistical description of data, Issues and Applications**.**

Unit-III	**[03]**	Max Marks:**04**

Data Preprocessing: Data cleaning, Data integration and Transformation, Data reduction, Data discretization and Concept Hierarchy Generation, Data mining primitives.

Unit-IV	**[03]**	Max Marks:**06**

Data Cube Technology: Efficient method for data cube computation, Cube materialization (Introduction to Full cube, Iceberg cube, Closed cube, Shell cube), General strategies for cube computation, Attribute oriented induction for data characterization, Mining class comparison, Discriminating between different classes.

Unit-V	**[05]**	Max Marks:**08**

Mining Frequent Patterns : Frequent patterns, Market basket analysis, Frequent itemsets, closed itemsets, association rules, Types of association rule (Single dimensional, multidimensional, multilevel, quantitative), Finding frequent itemset (Apriori algorithm, FP growth), Generating association rules from frequent itemset, Limitation and improving Apriori, From Association Mining to Correlation Analysis, Lift.

Unit-VI	**[07]**	Max Marks:**12**

Classification and Prediction: Definition (Classification, Prediction), Learning and testing of classification, Classification by decision tree induction, ID3 as attribute selection algorithm, Bayesian classification, Laplace smoothing, Classification by backpropagation, Rule based classifier (Decision tree to rules, rule coverage and accuracy, efficient of rule simplification)

Unit-VII	**[06]**	Max Marks:**10**

Cluster Analysis: Types of data in cluster analysis, Similarity and dissimilarity between objects, Clustering techniques: - Partitioning (k-means, k-means++, Mini-Batch k-means, k-medoids), Hierarchical (Agglomerative and Divisive), Density based (DBSCAN).

Text Book

1. Jiawei Han, Micheline Kamber and Jian Pei"Data Mining Concepts and Techniques", Third Edition, Elsevier, 2011. *ISBN-10. 0123814790* · ISBN-13. 978-0123814791

Reference Books

1. Alex Berson and Stephen J. Smith "Data Warehousing, Data Mining & OLAP", Tata McGraw – Hill Edition, Tenth Reprint 2007. *ISBN*: 1558604898
2. G. K. Gupta "Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006. ISBN-13: 978-8120350021 ISBN-10: 8120350022.
3. Pang-Ning Tan, Michael Steinbach and Vipin Kumar "Introduction to Data Mining", Pearson Education, 2007. ISBN 10: 1-292-02615-4, ISBN 13: 978-1-292-02615-2
4. K.P. Soman, Shyam Diwakar and V. Ajay "Insight into Data mining Theory and Practice", Easter Economy Edition, Prentice Hall of India, 2006. ISBN 10: 1-292-02615-4, ISBN 13: 978-1-292-02615-2

Course Outcome:

9. Learn and understand the usage and need of Data Warehouse
10. Learn various techniques for Data Warehouse and Data Mining.
11. Understand Market Basket Analysis

CourseCode : MDS-305	**Lab on Big Data (2) (P)**	*ClockHours:60* *Total Marks:***50**

Course Objectives:

1. To Describe the key issues in Big Data Management and experiment with the Hadoop framework
2. To Understand the structure and unstructured data by using NoSQL commands.
3. To Apply scientific computing algorithms for finding similar items and clustering.
4. To Develop problem solving and critical thinking skills in fundamental enable techniques like Hadoop & MapReduce.

Assignments

Practical 1: Installation of Hadoop Framework, it‘s components and study the HADOOP ecosystem
Practical 2: Write MapReduce program to find the grades of student’s
Practical 3: Write MapReduce program to Implementation of Matrix Multiplication using MapReduce.
Practical 4: MapReduce program to find the maximum temperature in each year
Practical 5: Write a program to implement bloom filtering
Practical 6: Write a program to implement K-means Clustering algorithm using MapReduce
Practical 7: Write a program to implement Twitter Sentiment Analysis System where we populate real-time sentiments for crisis management, service adjusting and target marketing using PYSpark
Practical 8: Install and Configure MongoDB to execute NoSQL Commands
Practical 9: HDFS: Setup a HDFS in a single node to multi node cluster, perform basic file system operation on it using commands provided, monitor cluster performance.
Practical 10: Hbase: Setup of Hbase in single node and distributed mode, write program to write into hbase and query it.
Practical 11: Elastic Search: Setup elastic search in single mode and distributed mode, define template, write data in it and finally query it.

Course Outcome:

Upon completion of the course, the students should be able to:

1. Configure Hadoop and perform File Management Tasks
2. Apply MapReduce programs to implement real life problems
3. Critically analyze huge data set using Hadoop distributed file systems and MapReduce
4. Apply different data processing tools like Pig, Hive and Spark
5. Critically analyze and apply structure and unstructured data by using NoSQL commands

CourseCode : MDS-**306**	**Lab on Advanced Machine Learning and Deep Learning (2) (P)**	*ClockHours:6***0** *Total Marks:***50**

Course Objectives:

1. To provide hands-on experience in implementing advanced machine learning and deep learning algorithms using modern tools and frameworks.
2. To enable students to analyze and preprocess data for designing and optimizing complex ML/DL models.
3. To develop skills in applying advanced techniques like ensemble learning, transfer learning, and GANs to solve real-world problems.
4. To instill proficiency in evaluating model performance and addressing challenges such as overfitting, bias, and scalability

Assignments

1. **Implementation of Ensemble Learning**
- Objective: Implement and analyze the performance of ensemble methods like Bagging, Boosting, and Random Forests.
- Task: Use datasets like UCI's Heart Disease or Titanic Survival to compare ensemble models with individual classifiers.

2. **Support Vector Machines (SVMs)**
- Objective: Develop a classification or regression model using SVMs with different kernels (linear, polynomial, RBF).
- Task: Use the Iris or Diabetes dataset to build and evaluate SVM models.

3. **Dimensionality Reduction Techniques**
- Objective: Apply PCA, t-SNE, and UMAP for feature reduction and visualization.
- Task: Visualize high-dimensional datasets (e.g., MNIST or CIFAR-10) in 2D space and analyze the results.

4. Implementation **of Autoencoders**
- Objective: Design and train an autoencoder for dimensionality reduction or denoising.
- Task: Use the MNIST dataset to create a basic autoencoder and evaluate its reconstruction quality.

5. Building **Convolutional Neural Networks (CNNs)**
- Objective: Design and train a CNN for image classification tasks.
- Task: Use the CIFAR-10 dataset to implement a custom CNN and analyze its performance.

6. Transfer **Learning**
- Objective: Use pre-trained models like VGG16, ResNet, or EfficientNet for image classification.
- Task: Fine-tune a pre-trained model on a custom dataset (e.g., Flowers or Fruits dataset).

7. Recurrent **Neural Networks (RNNs) and LSTMs**
- Objective: Implement RNNs and LSTMs for sequential data processing.
- Task: Train an LSTM model for sentiment analysis using the IMDB reviews dataset.

8. Transformer **Models**
- Objective: Experiment with transformer-based architectures for NLP tasks.
- Task: Use a transformer model like BERT or GPT to perform text classification or summarization.

9. **GANs (Generative Adversarial Networks)**
- Objective: Design and train a GAN for image generation.
- Task: Use the MNIST dataset to create a GAN and generate synthetic handwritten digits.

10. **Hyperparameter Tuning**
- Objective: Optimize ML/DL models using grid search, random search, or Bayesian optimization.
- Task: Experiment with hyperparameter tuning on models like Random Forests or CNNs.

11. **Time Series Forecasting**
- Objective: Build a deep learning model for time series prediction.
- Task: Use an RNN or LSTM model to forecast stock prices or weather data.

12. **Model Explainability (SHAP/LIME)**
- Objective: Analyze and interpret ML/DL models using explainability tools.
- Task: Apply SHAP or LIME to a Random Forest or Neural Network to understand feature importance.

13. **Object Detection with YOLO or Faster R-CNN**
- Objective: Implement an object detection model on image datasets.
- Task: Use a pre-trained YOLO or Faster R-CNN model to detect objects in a custom dataset.

14. **Reinforcement Learning (RL)**
- Objective: Explore RL by training an agent to solve an environment (e.g., OpenAI Gym).
- Task: Train an agent using Q-learning or DDPG to solve a cart-pole balancing task.

15. **Deploying a Deep Learning Model**
- Objective: Deploy a trained deep learning model as a web application or API.
- Task: Use Flask or FastAPI to create an API for a trained model and test it with real-world inputs.

16. **Self-Supervised Learning**
- Objective: Explore self-supervised techniques for label-scarce environments.
- Task: Train a contrastive learning model on CIFAR-10 and evaluate it on downstream tasks.

17. **Clustering using Advanced Techniques**
- Objective: Perform clustering using K-Means, DBSCAN, and Hierarchical Clustering.
- Task: Compare clustering algorithms on datasets like UCI Wine or Mall Customers.

18. **Image Segmentation**
- Objective: Implement image segmentation using U-Net or SegNet.
- Task: Train a model on datasets like the Oxford-IIIT Pet Dataset to segment objects.

19. **Multi-Task Learning**
- Objective: Build a model capable of solving multiple tasks simultaneously.
- Task: Use a dataset like CelebA to predict multiple attributes (e.g., gender, age, smile).

20. **Ethical Considerations and Bias Mitigation**
- Objective: Evaluate and mitigate biases in machine learning models.

- Task: Identify biases in models trained on datasets like COMPAS and implement mitigation strategies.

References Books:

1) Python for Data Science, Prof. Reghunathan Rengasamy, IIT Madras. (Free Available Online on NPTEL)
2) **Advanced Machine Learning with Python,** John Hearty, Packt Publishing, 2016, **ISBN**: 978-1784398637
3) **Python Machine Learning (3rd Edition),** Sebastian Raschka and Vahid Mirjalili, Packt Publishing, 2019, **ISBN**: 978-1789955750
4) **Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence,** Sebastian Raschka, Joshua Patterson, Corey Nolet, arXiv, 2020 https: //doi.org/10.48550 /arXiv.2002 .04803
5) **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow,** Aurélien Géron, O'Reilly Media, 2nd, 2019, *ISBN-10. 1492032646* · ISBN-13. 978-1492032649
6) **Python Machine Learning,** Sebastian Raschka, Vahid Mirjalili, Packt Publishing, 3rd Ed, 2019, ISBN: 9781789955750.

Course Outcome:

1. Implement advanced machine learning algorithms such as SVMs, ensemble methods, and clustering techniques to solve computational problems.
2. Design and train deep learning models like CNNs, RNNs, GANs, and transformers using frameworks like TensorFlow and PyTorch.
3. Optimize ML/DL models through hyperparameter tuning, transfer learning, and regularization techniques for enhanced performance.
4. Apply advanced ML/DL techniques to real-world applications in domains like computer vision, NLP, and time series analysis, while addressing ethical considerations.

Course Code: **MDSE-307A**	**Computer Vision (2) (T)**	*ClockHours:**30*** *Total Marks:**50***

Course Objectives:

1. Introduce the foundational concepts of computer vision.
2. Understand image processing and feature extraction techniques.
3. Explore applications of computer vision in object detection, recognition, and scene understanding

Unit-I	**[04]**	Max Marks:06

Introduction to Computer Vision: Overview of Computer Vision: History, Applications, and Challenges, Image Formation: Cameras, Lenses, and Projection Models, Basic Image Properties: Intensity, Color, and Geometry

Unit-II	**[06]**	Max Marks:**10**

Image Processing Basics : Image Enhancement: Histogram Equalization, Smoothing, and Sharpening, Edge Detection: Sobel, Canny, and Laplacian Operators, Morphological Operations: Erosion, Dilation, and Contour Detection.		
Unit-III	**[06]**	Max Marks:**10**
Feature Detection and Description: Corner Detection: Harris and Shi-Tomasi, Feature Descriptors: SIFT, SURF, ORB, Feature Matching: Brute-Force and FLANN-Based Matching, Applications: Object Matching and Motion Tracking.		
Unit-IV	**[06]**	Max Marks:**10**
Object Detection and Recognition: Template Matching, Basics of Object Detection using Haar Cascades, Introduction to Convolutional Neural Networks (CNNs) for Object Recognition, Applications in Face Detection and Scene Analysis.		
Unit-V	**[04]**	Max Marks:**07**
Motion Analysis and 3D Vision: Optical Flow: Lucas-Kanade and Horn-Schunck Methods, Structure from Motion (SfM): Basics of 3D Reconstruction, Stereo Vision: Depth Estimation.		
Unit-VI	**[04]**	Max Marks:**07**
Applications of Computer Vision: Image Segmentation for Semantic Understanding, Real-Time Applications: Augmented Reality, Autonomous Vehicles, and Surveillance		

References Books:

1) Szeliski, R. (2022). Computer vision: algorithms and applications. Springer Nature. ISBN-10. 3030343715 ; ISBN-13. 978-3030343712.
2) Richard Hartley and Andrew Zisserman (2004) .Multiple View Geometry in Computer Vision, 2nd edition , Cambridge University Press, ISBN:9780511811685
3) Shanmugamani, R. (2018). Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras. Packt Publishing ,Ltd. ISBN: 9781788295628.
4) Online Resources: OpenCV Documentation, PyTorch Tutorials.

Course Outcome:

1. By the end of this course, students will be able to:
2. Understand the foundational concepts of computer vision.
3. Understand image processing and feature extraction techniques:
4. Explore applications of computer vision in object detection, recognition, and scene understanding

*Course Code : M**DSE***-307B	**Lab on Computer Vision (2) (P)**	*ClockHours:6***0** *Total Marks:***50**

Course Objectives:

1. Gain hands-on experience with computer vision techniques and tools.
2. Implement and experiment with feature extraction, object detection, and motion analysis.
3. Use computer vision frameworks and libraries for real-world applications.

Assignments

1. Write a program to create 3D scene and experiment with different camera intrinsic parameters (focal length, sensor size).
2. Write a program to visualize the projection of 3D objects onto a 2D image plane using a basic camera model.
3. Load images and display basic properties such as pixel intensity, color channels, and geometry (e.g., width, height), Explore different color spaces (RGB, HSV) and manipulate images to highlight color properties.
4. Write a program to implement smoothing techniques (Gaussian blur) and sharpening filters on sample images.
5. Write a program to implement Harris and Shi-Tomasi corner detection algorithms, and detect corners in sample images.
6. Extract keypoints using SIFT, SURF, and ORB from sample images and visualize the keypoints on the images.
7. Use brute-force and FLANN-based methods for matching features between two images.
8. Write a program to implement object matching and motion tracking using features detected in video frames.
9. Use template matching to locate objects within an image. Experiment with different scaling and rotation parameters to test the robustness.
10. Write a program to implement face detection using pre-trained Haar cascade classifiers.
11. Use a pre-trained CNN (e.g., VGG16, MobileNet) for object classification tasks with a custom dataset or sample images.
12. Write a program to implement the Lucas-Kanade and Horn-Schunck methods for optical flow estimation on a video sequence and visualize the flow vectors.
13. Use a set of images taken from different viewpoints of a scene to implement basic 3D reconstruction.
14. Use a stereo pair of images to estimate depth information using stereo matching techniques.
15. Write a program to implement semantic segmentation using basic techniques like thresholding or more advanced methods like k-means clustering or watershed.

References Books:

1) Szeliski, R. (2022). Computer vision: algorithms and applications. Springer Nature. ISBN-10. 3030343715 ; ISBN-13. 978-3030343712.
2) Richard Hartley and Andrew Zisserman (2004) .Multiple View Geometry in Computer Vision, 2nd edition , Cambridge University Press, ISBN:9780511811685
3) Shanmugamani, R. (2018). Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras. Packt Publishing ,Ltd. ISBN: 9781788295628.
4) Online Resources: OpenCV Documentation, PyTorch Tutorials

Course Outcome:

1. On successful completion of the course, students will be able to:
2. Understand and apply the basics of computer vision in context of 3D graphics.

3. Use feature detection and feature description.
4. Apply knowledge for applications of computer vision in object detection, recognition, and scene understanding.

Course Code: **MDSE-308A**	**Large Language Model (2) (T)**	*ClockHours:***30** *Total Marks:***50**

Course Objectives:

1. To introduce the foundational concepts and architectures of Large Language Models (LLMs) and their applications.
2. To explore techniques for training, fine-tuning, and optimizing LLMs for diverse natural language processing tasks.
3. To understand the challenges of deploying LLMs, including computational efficiency, scalability, and ethical considerations.
4. To enable students to apply LLMs for real-world applications such as text generation, sentiment analysis, and question answering

Unit-I	**[04]**	Max Marks:**08**
Introduction to Deep Learning (Perceptron, ANN, backpropagation, CNN) Word vectors (Word2Vec, GloVe, fastText)		
Unit-II	**[06]**	Max Marks:**08**
Introduction to Statistical Language Models (N-gram LM, Perplexity, Smoothing) Language Models with CNN and RNN, Implementation of RNNs and LSTMs using PyTorch		
Unit-III	**[06]**	Max Marks:**08**
Introduction to Transformers, Sequence-to-sequence models, Beam search, Attention and self-attention, Positional Embedding, various tokenization strategies		
Unit-IV	**[06]**	Max Marks:**10**
Transfer Learning: ELMo, BERT (Encoder-only Model), Transfer Learning: GPT (Decoder-only Model), T5 (Encoder-decoder model), Prompting (hard and soft) and Instruction fine-tuning (FLAN), Advanced prompting (Chain of Thoughts, Graph of Thoughts, Prompt Chaining, etc.)		
Unit-V	**[04]**	Max Marks:**08**
Alignment with Human Feedback: RLHF, RLAIF, Parameter-efficient adaptation (Prompt tuning, Prefix tuning, LoRA)		
Unit-VI	**[04]**	Max Marks:**08**
Overview of recently popular models such as GPT4, Llama 3, Claude 3,Mistral, and Gemini Ethical NLP – Bias and Toxicity		

References Books:

1) "Large Language Models: A Deep Dive: Bridging Theory and Practice", Dr. Uday Kamath, Dr. Kevin Keenan, Dr. Garrett Somers, and Sarah Sorenson, Springer, 2024, ISBN-10: 3031656466, ISBN-13: 978-3031656460
2) "Build a Large Language Model (From Scratch)", Sebastian Raschka, Manning Publications, First Edition, 2024
3) "Understanding Large Language Models: Learning Their Underlying Concepts and Technologies", Thimira Amaratunga, 2024.

Course Outcome:

1. By the end of this course, students will be able to:
2. Understand the architecture and working principles of LLMs, including transformers, attention mechanisms, and pre-training methods.
3. Train and fine-tune LLMs like GPT, BERT, or T5 using modern frameworks and datasets.
4. Apply LLMs to solve advanced natural language processing problems, demonstrating practical proficiency in various tasks.
5. Evaluate the performance, ethical implications, and societal impact of LLMs, ensuring responsible and informed deployment.

*CourseCode : M**DSE**-**308B**	**Lab on Large Language Model (2) (P)**	*ClockHours:6***0** *Total Marks:***50**

Course Objectives:

1. To provide hands-on experience with training, fine-tuning, and deploying Large Language Models (LLMs) for real-world applications.
2. To explore the practical use of LLM libraries, frameworks, and datasets for solving advanced natural language processing (NLP) tasks.
3. To develop the ability to analyze the performance of LLMs and optimize them for efficiency and scalability.
4. To instill ethical and responsible practices in the implementation and application of LLMs

Assignments

1. **Introduction to Pre-trained LLMs**

- Objective: Explore and analyze pre-trained models (e.g., GPT, BERT, or T5) using libraries like Hugging Face.
- Task: Load a pre-trained model and perform simple text generation or fill-in-the-blank tasks.

2. **Text Classification with Fine-tuning**

- Objective: Fine-tune a pre-trained LLM for a text classification task (e.g., sentiment analysis or spam detection).
- Task: Use a labeled dataset (e.g., IMDB reviews) to fine-tune BERT or GPT and evaluate accuracy.

3. **Named Entity Recognition (NER)**

- Objective: Implement NER using a pre-trained LLM.
- Task: Fine-tune a model like BERT on a dataset such as CoNLL-2003 to identify entities (e.g., names, locations).

4. **Text Summarization**

- Objective: Generate text summaries using an LLM.
- Task: Use a model like T5 or BART to summarize large documents or articles and evaluate results with ROUGE scores.

5. **Custom Dataset Fine-tuning**

- Objective: Train an LLM on a custom dataset for a domain-specific application.

- Task: Collect or preprocess a dataset (e.g., product reviews, legal documents) and fine-tune a model.

6. **Zero-shot and Few-shot Learning**

- Objective: Explore the zero-shot and few-shot learning capabilities of LLMs.
- Task: Use GPT models to perform tasks like text classification or summarization without fine-tuning.

7. **Bias Detection and Mitigation**

- Objective: Analyze biases in LLMs and implement mitigation techniques.
- Task: Identify biases in generated text and apply fine-tuning or filtering techniques to reduce them.

8. **Prompt Engineering for LLMs**

- Objective: Experiment with different prompt designs to influence model outputs.
- Task: Create optimized prompts for tasks like story generation, coding help, or chatbot conversations.

9. **Chatbot Development**

- Objective: Build a conversational chatbot using an LLM.
- Task: Use GPT or a similar model to create a chatbot capable of multi-turn conversations on a specific topic.

References Books:

1) Natural Language Processing with Transformers: Building Language Applications with Hugging Face, Authors: Lewis Tunstall, Leandro von Werra, Thomas Wolf, Publisher: O'Reilly Media, Edition: 1st, 2022.
2) Deep Learning for Natural Language Processing: Creating Neural Networks with Python, Palash Goyal, Sumit Pandey, Karan Jain, Apress, Edition: 1st, 2018
3) Generative AI with Python and TensorFlow 2: Harness the Power of OpenAI's GPT-3 and Beyond, Joseph Babcock, Packt Publishing, Edition: 1st, 2021.
4) Advanced Natural Language Processing with TensorFlow 2, Ashish Bansal, Packt Publishing, Edition: 1st, 2021.
5) Applied Natural Language Processing with Python, Taweh Beysolow II, Apress, Edition: 1st, : 2018.
6) Pretrained Transformers for Text Ranking: BERT and Beyond, Jimmy Lin, Rodrygo L. T. Santos, Andrew Yates, Morgan & Claypool Publishers, Edition: 1st, 2021.
7) Building Large Language Models Applications: GPT, BERT, and Beyond, OpenAI Team Contributions *(unofficial),* Digital Library Collections *(hypothetical),* Edition: Beta

Course Outcome:

1. By the end of this laboratory course, students will be able to:
2. Fine-tune pre-trained LLMs like GPT, BERT, or T5 for specific NLP tasks using frameworks like Hugging Face, PyTorch, or TensorFlow.
3. Implement real-world NLP solutions such as text generation, summarization, translation, and sentiment analysis using LLMs.
4. Evaluate the performance of LLMs using metrics like BLEU, ROUGE, and perplexity and optimize them for resource efficiency.
5. Demonstrate an understanding of ethical considerations, biases, and limitations in deploying LLM-based solutions.

SEMESTER IV

Course Code: **MDS-401**	**Swayam Online (4) (T)**	*ClockHours:***60** *Total Marks:***100**
Major Courses MDS-303 must be completed from NPTEL /Swayam Online. The Subject will be given before the commencement of the semester		

Course Code: **MDS-402**	**Swayam Online (4) (T)**	*ClockHours:***60** *Total Marks:***100**
Major Courses MDS-303 must be completed from NPTEL /Swayam Online. The Subject will be given before the commencement of the semester		

Course Code: **MDS-403A / MDS-403B**	**Industrial Training (14) (P)** **OR** **Inhouse Research Project (14)(P)**	*ClockHours:***Semester** *Total Marks:****300***

Course Objectives:

1. To provide comprehensive learning platform to students where they can enhance their employ ability skills and become job ready along with real corporate exposure.
2. To enhance students' knowledge in a particular technology and to Increase self-confidence of students and helps in finding their own proficiency.
3. To cultivate student's leadership ability and responsibility to perform or execute the given task.

MDS- 403A: Industrial Training

Fourteen credits shall be awarded to the Industrial Training, which will commence in the IVth Semester and the final work and report will be completed at the end of IVth Semester of M.Sc. D.S. The student is expected to work on software development project. The project work should have coding part. Student will have to submit the bound project report in university prescribed format at the end of the semester. Student will have to appear for Project Viva-voce and the marks and the credits will be allotted at the end of IVth semester of M.Sc. D.S.

OR

MDS- 403A: Inhouse Research Project

Fourteen credits shall be awarded to the Inhouse Research Project, which will commence in the IVth Semester and the final work and report will be completed at the end of IVth Semester of M.Sc. D.S. The student is expected to work on software development project. The project work should have coding part. Student will have to submit the bound project report in university prescribed format at the end of the semester. Student will have to appear for Project Viva-voce and the marks and the credits will be allotted at the end of IVth semester of M.Sc. D.S.

Course Outcome:

1. Handle specialized technology and update themselves with latest changes in technological world with ability to communicate effectively.
2. Be multi-skilled IT professional with good technical knowledge, management, leadership and entrepreneurship skills.
3. Be able to identify, formulate and model problems and find engineering solution based on a systems approach.

*** * * * * * * * ***