LOYOLA ACADEMY

DEGREE & PG COLLEGE
OLD ALWAL, SECUNDERABAD - 500 010, TELANGANA, INDIA
(An Autonomous Institution Affiliated to Osmania University)
Re-accredited with 'A' Grade (III Cycle) by NAAC
A Catholic Christian Minority Institution (Co-Education)
A "College with Potential for Excellence" by UGC
www.loyolaacademy.edu.in **email:** info@loyolaacademy.edu.in **Ph:** 040-27862363/27860077

SYLLABUS

Of

M. Sc. Data Science

For the Academic Year 2026-2027

LOYOLA ACADEMY

DEGREE & PG COLLEGE
OLD ALWAL, SECUNDERABAD - 500 010, TELANGANA, INDIA
(An Autonomous Institution Affiliated to Osmania University)
Re-accredited with 'A' Grade (III Cycle) by NAAC
A Catholic Christian Minority Institution (Co-Education)
A "College with Potential for Excellence" by UGC
www.loyolaacademy.edu.in **email:** info@loyolaacademy.edu.in **Ph:** 040-27862363/27860077

SYLLABUS

of

M. Sc. Data Science

Principal **Dean** **Chairman, Board of Studies**

For the Academic Year 2026-2027 for the Batch 2026-2028

Index

Sl.No.	Content	Page No.
1	Programme Outcomes	1
2	Programme Specific Outcomes	2
3	Consolidated Scheme	3
4	Scheme of Instructions for Academic Year 2024-26	4
5	**FIRST SEMESTER SCHEME**	5
6	Communicative Competence (AEC-1)	6-7
7	Technical Seminar and Report Writing (SEC-1)	8
8	Mathematics for Data Science (DSC-1)	9-10
9	Artificial Intelligence (DSC-2)	11-12
10	Probability and Distribution Theory (DSC-3)	13-14
11	Python for Data Science (DSC-4)	15-16
12	Machine Learning (DSC-5)	17-19
13	Data Visualization using PowerBI Lab (Lab-I)	20-21
14	Python for Data Science Lab (Lab-II)	22
15	Machine Learning Lab (Lab-III)	23
16	**SECOND SEMESTER SCHEME**	24
17	Human Values and Professional Ethics (AEC-2)	25-26
18	Advanced Excel (SEC-2)	27
19	Design and Analysis of Algorithms |(GE-1)	28-30
20	Cloud Computing	31-32
21	Social Media Analytics	33-34
22	Web Analytics	35-36
23	Regression Analysis & Statistical Inference (DSC6)	37-39
24	Deep Learning (DSC7)	40-41
25	Advanced Databases (DSC8)	42-43
26	Data Analytics Lab (Lab-IV)	44-45
27	Deep Learning Lab (Lab-V)	46

28	Advanced Databases Lab (Lab-VI)	47
29	**THIRD SEMESTER SCHEME**	48
30	Soft Skills (AEC-3)	49-50
31	Mini Project (SEC-3)	51
32	Time Series Analysis and Forecasting Techniques (DSC9)	52-53
33	Big Data Security	54-55
34	Reinforcement Learning	56-57
35	Scalable Architectures for ML Applications	58-59
36	Internet Of Things (DSC10)	60-61
37	Deep Learning (DSC11)	62-63
38	Natural Language Processing (DSC12)	64-65
39	Internet Of Things Lab (Lab VII)	66
40	Deep Learning Lab (Lab VIII)	67
41	Natural Language Processing Lab (Lab –IX)	68
42	**FOURTH SEMESTER SCHEME**	69
43	Project	70
44	BOS Signing Sheet	
45	BOS Agenda	
46	BOS Minutes	

M.SC. DATA SCIENCE

PROGRAMME OUTCOMES

PO1: Engage in continuous reflective learning in the context of technology and scientific achievement.

PO2: Identify the need and scope for the Interdisciplinary area.

PO3: Understand the Professional, ethical and social responsibilities.

PO4: Enhance disciplinary competency, employability and technical skills.

DEPARTMENT OF M.SC. DATA SCIENCE

PROGRAMME SPECIFIC OUTCOMES

PSO1: Abstract Thinking: Ability to understand the abstract concepts that lead to various data science theories in mathematics, statistics and computer science.

PSO2: Problem Analysis and Design Ability: To identify, analyze and design solutions for data science problems using fundamental principles of mathematics statistics and relevant domain disciplines.

PSO3: Modern software tool usage: Acquire the skills in handling data science programming tools towards problem solving and solution analysis for domain specific problems.

PSO4: Societal and Environmental Concern Utilize the data science theories for societal and environmental concerns

PSO5: Professional Ethics: Understand and commit to professional ethics and cyber regulation.

Loyola Academy, Old Alwal, Secunderabad-500 010

Loyola Academy Degree and PG College
M.Sc. Data Science (2026-27) of Batch 2026-28(Choice Based Credit System)

Year	Sem	Course-1	Course-2	Course-3	Course-4	Course-5	Course-6	Course-7		Hours	Credits	
I	I	*Communicate Competence(2) (AEC-1)	*Technical Seminar and Report Writing(2) (SEC-1)	Mathematics for Data Science (4) (DSC1)	Artificial Intelligence (4) (DSC2)	Probability & Distribution Theory, Data Visualization using PowerBI Lab (4+2) (DSC3)	Python for Data science (4+2) (DSC4)	Machine Learning (4+2) (DSC5)	*PLANET (Community Work)	36	26	AEC=1 SEC=1 DSC=5 PLANET =1
I	II	*Human Values and Professional Ethics(2) (AEC-2)	*Advanced Excel(2) (SEC-2)	(GE-InterDisciplinary -1) Design and Analysis of Algorithms (4)	Elective-I Cloud Computing / Social Media Analytics/Web Analytics (DSE-I)(4)	Regression Analysis & Statistical Inference (4+2) (DSC6)	Advanced Databases (4+2) (DSC7)	Deep Learning (4+2) (DSC8)	36		26	AEC=1 SEC=1 ID=1 DSE=1 DSC=3
II	III	*Soft Skills(Interview Skills, Group discussion) (2) (AEC-3)	Mini Project Seminar (2) (SEC-3)	Agentic AI(4) (DSC9)	Elective-II Quantum Computing/ Reinforcement Learning /Scalable Architectures for ML Applications (DSE-II) (4)	Big Data Analytics Through Hadoop (4+2) (DSC10)	Natural Language Processing (4+2) (DSC11)	Computer Vision (4+2) (DSC12)	36		28	AEC=1 SEC=1 DSE=1 DSC=4
II	IV		Dissertation (15 weeks)						36		12	
			Legend: 1. Ability Enhancement Course (AEC) : 03 2. Interdisciplinary Elective : 01 3. Skill Enhancement Course (SEC) : 03 4. Disciple Specific Core : 12 5. Discipline-Specific Elective (DSE) : 02 6. Planet (Community Work) :01 *AEC is excluded for SGPA and CGPA calculation but it is compulsory to pass in the examination. Planet Program, 1 credit added in 1st semester(Non-CGPA), #At the end of Second Semester students will take a one-month summer project of 2 credits.							Total	**144**	**92**	

Categories of Courses and Credits

Abbreviation	Full Form	Credits	Abbreviation	Full Form	Credits
AEC	Ability Enhancement Course	06*	Practical	Practical Core Course	18
ID	Interdisciplinary Elective	04	Dissertation	Dissertation	12
SEC	Skill Enhancement Course	04* 2			
DSE	Discipline-Specific Elective	08			
PLANET	*PLANET (Community Work	01*			
DSC	Disciple Specific Core	48			
Total Credits		**92**			

SCHEME OF INSTRUCTION FOR M.Sc. DATA SCIENCE (2024-26)

I YEAR (2026-27) OF 2026-28 BATCH (CBCS)					
SEMESTER-I			**SEMESTER-II**		
THEORY			**THEORY**		
1	MDS24101	Communicative Competence (AECC-1)	1	*MDS24201	* Human Values & Prof Ethics (AEC-2)
2	MDS20102	Technical Seminar and Report Writing (SEC-1)	2	*MDS26202	*Advanced Excel(SEC-2)
			3	MDS26203	Design and Analysis of Algorithms |(GE-InterDisciplinary-1)
CORE			**CORE**		
3	MDS24103	Mathematics for Data Science (DSC-1)	4	MDS20204	Elective-I **(DSE-1)**
4	MDS26104	Artificial Intelligence(DSC-2)		MDS22204A	Cloud Computing
5	MDS24105	Probability and Distribution Theory (DSC-3)		MDS22204B	Social Media Analytics
6	MDS22106	Python for Data Science (DSC-4)		MDS24204C	Web Analytics
7	MDS26107	Machine Learning (DSC-5)	5	MDS24205	Regression Analysis & Statistical Inference (DSC6)
			6	MDS26206	Deep Learning (DSC7)
			7	MDS26207	Advanced Databases (DSC8)
PRACTICALS			**PRACTICALS**		
8	MDS26108	Data Visualization using Power BI Lab (Lab-I)	8	MDS20208	Data Analytics Lab (Lab-IV)
9	MDS22109	Python for Data Science Lab (Lab-II)	9	MDS26209	Deep Learning Lab (Lab-V)
10	MDS26110	Machine Learning Lab (Lab-III)	10	MDS26210	Advanced Databases Lab (Lab-VI)

II YEAR (2025-26) OF 2024-26 BATCH (CBCS)					
SEMESTER-III				**SEMESTER-IV**	
THEORY				**THEORY**	
1	MDS21301	Soft Skills (Interview Skills, Group discussion) **(AECC-3)**	1	MDS21401	Major Project
2	MDS21302	Mini Project **(SEC-3)**			
CORE					
3	MDS25303	Time Series & Forecasting Techniques (DSC**9**)			
4	MDS23304	Elective-II **(DSE-II)**			
	MDS23304A	Big Data Security			
	MDS23304B	Reinforcement Learning			
	MDS23304C	Scalable Architectures for ML Applications			
5	MDS25305	Internet of Things **(**DSC**10)**			
6	MDS25306	Deep Learning **(**DSC**11)**			
7	MDS25307	Natural Language Processing**(**DSC**12)**			
PRACTICALS					
8	MDS21308	Internet of Things Lab **(Lab-VII)**			
9	MDS23309	Deep Learning Lab **(Lab-VIII)**			
10	MDS25310	Natural Language Processing Lab **(Lab-IX)**			

FIRST SEMESTER SCHEME

YEAR-WISE AND SEMESTER-WISE DISTRIBUTION OF SUBJECTS M.Sc. DATA SCIENCE
FIRSY SEMESTER
ACADEMIC YEAR 2026-27 OF 2026-28 BATCH (CBCS)

Sl. No.	Subject Code	Title of the Subject	Hours /Week	Credits	Marks			Duration of Exam (hrs.)
					Internal	External	Total	
1	*MDS24101	*Communicative Competence **(AEC-1)**	2	2*	20	30	50	2
2	*MDS20102	*Technical Seminar and Report Writing **(SEC-1)**	2	2*	50	-	50	-
DSC								
3	MDS24103	Mathematics for Data Science **(DSC-1)**	4	4	40	60	100	3
4	MDS26104	Artificial Intelligence**(DSC-2)**	4	4	40	60	100	3
5	MDS24105	Probability and Distribution Theory **(DSC-3)**	4	4	40	60	100	3
6	MDS22106	Python for Data Science **(DSC-4)**	4	4	40	60	100	3
7	MDS26107	Machine Learning **(DSC-5)**	4	4	40	60	100	3
PRACTICALS								
8	MDS26108	Data Visualization using Power BI Lab **(Lab-I)**	4	2	40	60	100	3
9	MDS22109	Python for Data Science Lab **(Lab-II)**	4	2	40	60	100	3
10	MDS26110	Machine Learning Lab **(Lab-III)**	4	2	40	60	100	3
11		*PLANET (Community Work)		1*				
Total			**36**	**26**	**390**	**510**	**900**	

*** Ability Enhancement Course (AEC) and Skill Enhancement Course (SEC) is excluded for SGPA and CGPA calculation, but it is compulsory to pass the examination.**

CIA components for Internal marks of theory paper: i) Assignment: 3M ii) Attendance: 4M iii) Group Discussion OR Presentation OR Case Study OR Computer Based Test: 5M iv) Mid Exam 1: 12.5M v) Mid Exam 2: 12.5M vi) Seminar or Viva Voce: 3M

CIA components for Practical Internal marks: i) Attendance: 5M ii) Observation/record book: 5M iii) Practical skill acquired: 10M iv) Pre final: 20M.

*Ability Enhancement Compulsory Course (AEC) * Discipline Specific Core (DSC)

COMMUNICATIVE COMPETENCE

Credits: 2 **Semester: 1**
Course Code: MDS24101 **No of Lecture Hours: 30**

Course Objective:

- English language course will help students to develop a natural and accurate style of English pronunciation.
- English language course will also improve your ability to express ideas clearly and confidently in English.

Course Outcome:

CO1: Understand the features of Indian English and its differences from formal expressions and apply this knowledge in both written and verbal communication. (Knowledge/Comprehension)
CO2: Develop effective verbal and non-verbal communication skills, including rate of speech, pitch, tone and clarity of voice, to communicate with different audiences. (Application/Analysis)
CO3: Demonstrate presentation skills, including the ability to organize and deliver presentations in different formats, such as informative, persuasive and demonstrate presentations. (Application/Analysis)
CO4: Apply the skills and techniques required for formal and informal interviews, including creating the right ambience, effective questioning and active listening for successful job interviews and group of discussions. (Application/Analysis)
CO5: Develop proficiency in functional grammar, including the correct usage of tenses, phonetics, articles, sentence correction and subject-verb concord to enhance written and oral communication skills. (Synthesis/Evaluation)

Course Content	Hours Allotted	Pedagogy
Module I: • Features of Indian English • Informal conversation Vs Formal expression • Verbal and non-verbal communication • Barriers to effective communication – kinesics	**(6)** 2 1 1 2	Power point presentations / Lectures
Module II: • What is Communicative competence- Oral, aural, Writing, Reading • Rate of speech, pitch, tone- Clarity of voice • Presentation Skills • Types of Presentation	**(6)** 2 2 2 2	Case Studies / Lectures
Module III: • Formal and informal interviews • Ambiance and polemics • Interviewing skills and types of interviews • Group discussions, • Curriculum vitae	**(6)** 1 1 2 1 1	Case Studies

Module IV: • Tenses • Articles • Correction of Sentences • Subject Verb Concord	**(6)** 2 1 1 2	Lectures
Module V: • Letter- writing business letter • Analysis of sample letters collected from industry- email, fax • Writing a research paper	**(6)** 2 2 2	Case Studies

Learning Resources	
1.	Rajendra Pal, J S KorlahaHi, **Essentials of Business Communication**, Sultan Chand & Sons, New Delhi.
2.	Andrea J. Rutherford, **Basic Communication Skills for Technology**, Pearson Education Asia, Patparganj, New Delhi-92.
3.	V. Prasad, **Advanced Communication Skills**, Atma Ram Publications, New Delhi.
4.	Raymond V. Lesikav; John D. Pettit Jr.; **Business Communication; Theory & Application**, All India Traveller Bookseller, New Delhi-51.
5.	R. K. Madhukar, **Business Communication**, Vikas Publishing House Pvt. Ltd.

TECHNICAL SEMINAR AND REPORT WRITING

Credits: 2 **Semester: 1**
Course Code: MDS20102 **No of Lecture Hours: 30**

Course Objective:

- Awareness of how to use values in improving your own professionalism.
- Learning about personal and communication styles for personal building.
- Learning values management.
- Latest Research awareness in the interested domain

Course Outcome:

- The overall purpose or goal from participation in an educational activity for personal development and current trends in research

Course Content	Hours Allotted	Pedagogy
1. Oral presentation is an important aspect of higher education. The objective of the seminar is to prepare the student for systematic independent study of the 2. art topics in the broad area of his/her specialization. 3. Seminar topics can be chosen by the students with the advice from the faculty members. 4. Students are exposed to the following aspects of seminar presentations. • Literature survey • Organization of the material • PPT Presentation • Technical writing	30	Case Studies / Review of research articles/Lectures
Each student is required to Submit one page of Synopsis of the seminar talk two days before for display on notice board. 1. Give 20 minutes PPT presentation, followed by 10 minutes discussion. 2. Submit a report on the seminar topic with a list of references and slides used within a week. Seminars are to be scheduled in the 5th week of the semester. The sessional marks will be awarded to the students by at least 2 faculty members on the basis of an oral and written presentation as well as their involvement in the discussion.		Power point presentations

MATHEMATICS FOR DATA SCIENCE

Credits: 4 **Semester: 1**
Course Code: MDS24103 **No of Lecture Hours: 60**
Prerequisite: Basic Mathematics

Course Objective:

- The objective of this course is to introduce the basic notions of vector spaces, linear algebra and their applications of linear algebra.

Course Outcome:

CO1: Construct mathematical arguments that relate to the study of introductory linear algebra.
CO2: Determine whether a square matrix is diagonalizable and compute its diagonalization.
CO3: Compute the eigenvalues and eigenvectors of a square matrix and determine the dimension of the corresponding eigenspaces.
CO4: Analyse finite and infinite dimensional vector spaces and subspaces over a field and their properties, including the basis structure of vector spaces. Use the definition and properties of linear transformations and matrices of linear transformations.
CO5: Explain orthogonality on vector spaces and compute inner products and including Gram-Schmidt Orthogonalization.

Course Content	Hours Allotted	Pedagogy
Module I: Linear Equations in Linear Algebra • Trace and Rank of a Matrix and their properties. • Systems of linear equation-Row reduction and Echelon forms. • Matrix operations-The inverse of a matrix-Characterizations of Invertible Matrices. • Solution set of linear systems of equation Ax=0 and Ax=b.	**(12)** 3 3 2 4	Power point presentations / Lectures
Module II: **Eigenvalue Problems** • Charactertistic roots and vectors. • Cayley-Hamilton theorem (No proof)- Problems. • Diagonalization of a Matrix. **Quadratic forms** • Real quadratic forms. • Constrained Optimization. • The Singular value decomposition. • Index and signature	**(12)** 2 2 2 2 2 1 1	Case Studies / Review of research articles
Module III: Vector Spaces • General properties of Vector Space. • Vector Subspaces - Algebra of subspaces. • Linear Span- Linear sum of two subspaces. • Linear Dependence and Linear Independence of vectors. • Basis of a vector Space - Finite Dimensional Vector Space.	**(12)** 2 2 2 3 3	Assignments
Module IV: Linear Transformations	**(12)**	Lectures

• Linear Transformations – Linear Operators. Range and Null Spaces of a Linear Transformation. • Rank and Nullity of Linear Transformation. • Matrix of a Linear Transformation.	4 4 4	
Module V: Inner-product spaces • Inner product,length and orthogonality. • Orthogonal sets and bases. • Orthogonal projections- • The Gram-Schmidt Orthogonalization Process	**(12)** 3 3 3 3	Assignments / Lectures

Learning Resources	
1.	"Linear Algebra and its Applications" by David C. Lay, Steven R. Lay, Judi.J. Mcdonald, Fifth Ed., 2016 Pearson.
2.	Matrices by AR. Vasishtha Krishna Prakashan Mandir, Meerut.
3.	Gilbert Strang (2026): Introduction to linear algebra 5/e., Wellesley-Cambridge.
4.	Rao, C.R. Mith4 S.K. (1971): Generalized inverse of matrices and its applications, John Wiley & Sons Inc.
	RELATED ONLINE CONTENT [MOOC, SWAYAM, NPTEL, Websites etc.] https://nptel.ac.in/courses/111104137 2. https://nptel.ac.in/courses/111106051 3. https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/

ARTIFICIAL INTELLIGENCE

Credits: 4 **Semester: I**

Course Code: MDS26104 **No of Lecture Hours: 60**

Prerequisite: Basic Statistics and Programming

Course Objective:

- To understand some of the topics of AI such as Problem Space, Knowledge Representation, Searching Techniques, Constraint Satisfaction and planning.

Course Outcome:

CO1: Defines Artificial Intelligence, Agents and Problem-Solving Agents.

CO2: Explains about Knowledge Representation, Forward and Backward Chaining.

CO3: Differentiate between various Searching Techniques.

CO4: Apply various game playing techniques.

CO5: Explain about Heuristic, Forward and Backward State Space Search.

Course Content	Hours Allotted	Pedagogy
Module I: Title: Introduction ● AI Problems, Intelligent agents: Agents and Environments, Intelligent agents: ● Agents and Environments. ● Concept of Rationality, the Nature of Environments. ● Structure of Agents, Problem Solving Agents. ● Problem Formulation.	**(12)** 4 3 3 2	Powerpoint presentations / Lectures
Module II: Title: Knowledge Representation & Reasons ● Knowledge Representation: Introduction, Approaches to Knowledge Representation ● Knowlegde- Based Agents, the Wumpus World. ● Propositional Logic: Reasoning patterns in propositional logic- Resolution. ● Forward & Backward Chaining. ● Inference in First Order Logic: Propositional vs First Order Inference.	**(12)** 2 2 4 4	Case Studies / Review of research articles
Module III: Title: Searching ● Searching for solutions, Applications of SearchAlgorithms. uniformed search strategies-Depth Limited Search, Dijkstra Algorithm, ● Bi-direction Search, Comparing Uninformed Search Strategies. ● Search with Partial Information (Heuristic Search), TSP Problem. ● Best First Search, A* Search, Hill Climbing, Simulated Annealing Search.	**(12)** 3 3 3 3	Assignments
Module IV: Title: Constraint Satisfaction Problem ● Backtracking search for CSP's local search for Constraint Satisfaction Problem. ● Game Playing: Games, Min-Max Algorithm.	**(12)** 3	Lectures

● Alpha-Beta Pruning. **Planning** ● Classical Planning Problem, Language of Planning Problem. ● Planning with State Space Search. ● Forward State Space Search, Backward State Space Search. ● Heuristics for state space search.	3 3 3	
Module V: Title: **Expert Systems** ● Architecture of Expert Systems, Role of Expert System. **Generative AI** ● Introduction to GEN AI ● Overview of Generative Models ● Discriminative vs. Generative Models. ● Strengths and Weaknesses. ● Applications.	**(12)** 2 2 2 2 2 2 2	Assignments / Lectures

LearningResources	
1.	Stuart Russel, Peter Norvig, (2009), Artificial Intelligence- A modern Approach, 3rd Edition, Pearson Education.
2.	Saroj Kaushik, Artificial intelligence, Cengage Learning India, First Edition, 2011.
3.	E. Rich and K. Knight, (2008), Artificial Intelligence, 3rd Edition, Tata McGraw Hill.
4.	Patterson, (2009), Artificial Intelligence and Expert Systems, 2nd Edition, PHI.
5.	Giarratana / Riley, (2004), Expert Systems: Principles and Programming, 4th Edition, Thomson.
6.	ZonunfeliRalte, Indrajit Kar, , “Learn Python Generative AI, BPB publications, 2024. (Unit-V) (Cost- Rs 782)*
7.	Karthikeyan Sabesan, Sivagamisundari, Nilip Dutta, , Generative AI for Everyone, BPB publications, 2025. (Cost- Rs 999)*

PROBABILITY AND DISTRIBUTION THEORY

Credits: 4 **Semester: I**
Course Code: MDS24105 **No of Lecture Hours: 60**
Prerequisites: Basics in Mathematics and Statistics

Course Objective:

- To incorporate the concepts of probability theory and its applications as the core material in building theoretical ideas along with the practical notion. To integrate the intrinsic ideas of preliminary and advanced distributions to correlate with the real-world scenarios.

Course Outcome:

CO1: Develop problem-solving techniques needed to calculate probability and conditional probability.
CO2: Formulate fundamental probability distribution functions from random variables and calculate the expectations and generating functions with properties.
CO3: Describe commonly used univariate discrete probability distributions.
CO4: Describe commonly used univariate continuous probability distributions.
CO5: Apply sampling distributions to test the hypothesis.

Course Content	Hours Allotted	Pedagogy
Module I: Probability:	**(12)**	Power point presentations / Lectures
• Introduction to probability - random experiment, sample space, random events, algebra of events, equally likely outcomes.	2	
• Definition of probability, probability axioms.	1	
• Addition theorem on probability (theorem with proof and problems).	2	
• Bonferroni's inequality, Boole's inequality,	2	
• Conditional probability Multiplication theorem on probability (theorem with proof and problems)	2	
• Independence - pairwise and mutually independent events.	1	
• Total probability and Baye's law (theorem with proof and problems).	2	
Module II: Title:	**(12)**	Case Studies / Review of research articles
• **Random Variables** – discrete and continuous random variables. Distribution functions - probability mass function, probability density function, cumulative distribution functions and their properties.	3	
• **Bivariate random variables (theory and problems - no derivations)**: Joint probability functions, marginal probability functions, Conditional probability functions, independence of random variables. Transformation techniques: One variable transformation and many variables transformation (theory and problems - no derivations)	3	
• **Expectations:** expectation, variance, and covariance of random variables. Conditional expectation and conditional	3	

variance. • **Generating functions (Only problems - No derivations):** Moment generating function (MGF), cumulant generating function (KGF), Characteristic germinating function (CGF), Probability generating functions (PGF) • **Inequalities (only statements, no theorems):** Markov, Jensen's, and Chebyshev's inequalities. Weak law of large numbers, strong law of large numbers and Kolmogorov theorem (statement only) Central limit theorem (statement only).	2 1	
Module III: Discrete distributions (No derivations only problems): • **Definition, properties, and applications with numerical problems of the following distribution functions:** i. Bernoulli ii. Binomial iii. Poisson, Geometric iv. Hypergeometric v. Negative Binomial vi. Multinomial distributions vii. Discrete Uniform distribution	**(12)** 8 3 1	Assignments
Module IV: Continuous distributions (No derivations only problems) • **Definition, properties, and applications with numerical problems of the following distribution functions:** i. Uniform ii. Normal distribution function iii. Exponential iv. Gamma v. Beta distributions (First and Second kind) vi. Weibull vii. Cauchy viii. Laplace distributions ix. lognormal, x. logistic xi. Pareto xii. Rayleigh distribution	**(12)** 5 3 2 2	Lectures
Module V: Sampling distributions (No derivations) • Sampling distribution of mean, proportions, sample variance. • **Definition, properties, and applications of the following distribution functions:** i. Chi-square distribution ii. t distribution iii. F distribution order statistics (definition).	**(12)** 3 8 1	Assignments / Lectures

Learning Resources	
1.	Sheldon Ross; A First Course in Probability, Pearson, 2014.
2.	Parimal Mukhopadhyay; An Introduction to the Theory of Probability, World scientific, 2012
3.	Irwin Miller, Marylee's Miller, John E. Freund's; Mathematical Statistics, Pearson, 2017.
4.	FetsjeBijma, Marianne Jonker and Aad van der Vaart; Introduction to Mathematical Statistics, Amsterdam University Press, 2018.
5.	Krishnamoorthy, K., Handbook of Statistical Distributions with Applications, Chapman & Hall/CRC, 2006.
6.	Rohatgi, V.K. and Ebsanes Saleh, A.K. Md., An introduction to Probability and Statistics, 2nd Ed., John Wiley & Sons, 2002.
7.	Shanmugam, R., Chattamvelli, R. Statistics for scientists and engineers, John Wiley, 2015.

PYTHON FOR DATA SCIENCE

Credits: 2 **Semester: I**
Course Code: MDS22106 **No of Lecture Hours: 60**
Prerequisites: Basics of Computers and Programming

Course Objective:
- To learn the basic concepts and usage of variables, expressions and practice the use of functions in Python programming language.
- To identify and practice different conditionals and implement recursive functions.
- To understand the concepts of strings, lists, and dictionaries, practice the use of classes methods, overloading, and polymorphism.
- To learn different usage of various packages

Course Outcome:
CO1: Apply the concept of objects, variables, control structures, and numbers.
CO2: Apply the concept of data structures for python.
CO3: Explain Functions, Modules, and Packages.
CO4: Explain Files, Exceptions and OOPs concepts
CO5: Apply the concepts of GUI programming

Course Content	Hours Allotted	Pedagogy
Module I:	**(12)**	Power point presentations / Lectures
• Python Basics: Origin, Features, Statements and Syntax, Variable Assignment, Identifiers, Basic Style Guidelines, Memory Management.	2	
• Python Objects: Python Objects, Type Object, Null Object, Object Identity Comparison, type () and isinstance (), Categorizing the Standard Types.	2	
• Conditionals and Loops: if, else, elif statement, Conditional Expressions, while, for statement, range() Built-in Function, break, continue, pass.	4	
• Numbers: Introduction, Integers, Double Precision Floating Point Numbers, Complex Numbers, Operators, Built-in and Factory Functions, Other Numeric Types	4	
Module II:	**(12)**	Case Studies / Review of research articles
• Sequences: Strings, Lists and Tuples- Sequences, Strings, Strings and Operators, String-Only Operators, Built-in Functions, String Built-in Methods	4	
• Lists: Introduction, Operators, Built-in Functions, List Type Built-in methods. Tuples- Introduction, Tuple Operators and Built-in Functions	3	
• Dictionaries: Introduction, Mapping Type Operators, Mapping Type Built-in and Factory Functions, Mapping Type Built-in Methods, Dictionary Keys.	3	
• Set Types: Introduction, Set Type Operators, Built-in Functions and Methods	2	

Module III: • Functions: Introduction, Calling Functions, Creating Functions, Return values, parameters, local and global scope, Function composition, recursion. • Modules: Modules, Namespaces, Importing Modules, Features of Module Import, Built-in Functions • Packages: NumPy, Pandas, Matplotlib	**(12)** 4 2 6	Assignments
Module IV: • Files: Text files, reading and writing files, format operator, Command line arguments, errors and exceptions, Modules. • Exception: Introduction, Exceptions in Python, Detecting and Handling Exceptions, with Statement, raise Statement, Assertions-assert Statement, Standard Exceptions, Creating Exceptions. • Object Oriented Programming: Classes and Objects, Classes and Functions, Classes and Methods, Working with instances, Sub-classing and Derivation, Inheritance, Multiple Inheritance, BIF for Classes, Instances and Other Objects.	**(12)** 3 4 5	Lectures
Module V: • GUI Programming: Tkinter and Python programming, Tkinter examples- Label, Button, Checkbutton, Entry, Canvas, Listbox, Message, Radiobutton, Scale • GUI Programming: Menu, Menubutton, Scrollbar, Spinbox, Slider, Dialogs, Text, tkMessageBox, PanedWindow, TopLevel, Frame	**(12)** 6 6	Assignments / Lectures

Learning Resources	
1.	Chun, Wesley J. Core Python Programming. 2nd Edition. USA: Pearson Education (Unit I, II, III, IV)
2.	Chun, Wesley J. Core Python Applications Programming. 3rd Edition. USA: Pearson Education (Unit V)
3.	Allen B. Downey Think Python: How to Think like a Computer Scientist. 2nd edition, Updated for Python 3, Shroff/O 'Reilly Publishers, 2016.

MACHINE LEARNING

Credits:4 **Semester: I**
Course Code: MDS26107 **No of Lecture Hours: 60**
Prerequisites: Linear Algebra, Probability, Python

Course Objective:

To know data and preprocessing
To learn machine learning concepts
To design a learning system and know about regression tasks
To design a learning system and know about the classification tasks
To examine the concepts of unsupervised learning

Course Outcome:
CO1: Define data and pre-processing problems.
CO2: Describe machine learning concepts
CO3: Illustrate regression tasks
CO4: Illustrate classification tasks
CO5: Explain unsupervised learning concepts

Course Content	Hours Allotted	Pedagogy
Module I: Title: **Introduction to Data and Data Pre-processing** • Data, Basic Data Types - Nominal, Ordinal, interval, Ratio scaled • Data Cleaning - Missing Values, Noisy Data, Outlier Detection, Correction and Removal, • Data Cleaning as a Process, Data Imputation Techniques • Data Integration, Data Transformation, • Data Reduction - Attribute Subset Selection, Numerosity Reduction. • Association Rule Learning - Apriori, Eclat	**(12)** 2 2 2 2 2 2	Power point presentations / Lectures
Module II: Introduction to Machine Learning: • Overview-Working, Features, Need for Machine Learning, History • Artificial Intelligence verses Machine Learning • Applications of Machine Learning • Machine Learning Lifecycle • Design of a Learning System • Types of Machine Learning • Issues in Machine Learning	**(12)** 3 1 1 2 2 2 1	Case Studies / Review of research articles
Module III: Supervised Learning - I: • Simple Linear Regression • Multiple Linear Regression, Polynomial Regression • Ridge Regression • Lasso Regression	**(12)** 2 2 2 2	Assignments

• Evaluating Regression Models, Model Selection • Ensemble Methods-Bagging, Boosting	2 2	
• **Module IV: Supervised Learning – II:** • Classification, Classification verses Regression • Logistic Regression • K-Nearest Neighbor (KNN) • Support Vector Machine Regression and Classification • Decision Tree Regression and Classification • Random Forest Regression and Classification • Naïve Bayes Classifier • Evaluating Classification Models	**(12)** 2 1 1 2 2 2 1 1	Lectures
Module V: Unsupervised Learning: • Clustering • K-Means Clustering • Expect-ation-Maxi-mization Algorithm • Hierarchical Clustering Methods • Density-Based Clustering • Dimensionality Reduction-PCA • Collaborative Filtering-Recommendation systems	**(12)** **2** **2** 1 2 1 2 2	Assignments / Lectures

Learning Resources	
1.	Jiawei Han, Micheline Kamber, Jian Pei (2012), Data Mining: Concepts and Techniques, 3rd Edition, Elsevier, United States of America. India.
2.	Christopher Bishop, "Pattern Recognition and Machine Learning" Springer, 2007.
3.	Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012.
4.	Ethem Alpaydin, “Introduction to Machine Learning", MIT Press, Third Edition, 2014.
5.	Tom Mitchell, "Machine Learning", McGraw-Hill, 1997.

DATA VISUALIZATION USING POWERBI LAB (LAB-1)

Credits: 2 **Semester: I**
Course Code: MDS26108 **No of Lecture Hours: 60**

Course Objective:
- To Acquire the skill of Data Visualization through POWERBI

Course Outcome:
CO1: Import, clean, and transform structured and semi-structured datasets using Power BI Query Editor for analytical purposes.
CO2: Integrate multiple data sources and implement data modelling techniques to ensure accuracy and performance of reports.
CO3: Apply DAX functions to create calculated columns, measures, and advanced analytics for meaningful insights.
CO4: Design and develop interactive reports using appropriate charts, KPIs, slicers, and filters to represent complex data effectively.
CO5: Evaluate visualization effectiveness and optimize dashboards based on usability, clarity, and decision-making requirements.

Lab Experiments:
1. Get started creating in the Power BI service (Installation & Setup).
2. Program to implement how to add Data from different data sources.
3. Program to create KPI Visuals from Scratch.
4. Program to demonstrate Create Report visuals.
5. Program to create a report from an Excel file in the Power BI service.
6. Program to customize Tooltips in Power BI.
7. Program to add a filter to a report in Power BI.
8. Program to create a slicer.
9. Program to create a Field Parameter.
10. Program to create Buttons in Reports.
11. Program to create a Report Bookmarks.
12. Program to apply a Built-in Report Theme.
13. Program to create a Drill through Page.
14. Program to implement using Grouping.
15. Program to implement using Binning.
16. Program to enable Gridlines, Snap-to-grid, and Smart Guides.
17. Program to apply Page display settings in a Power BI Report.
18. Program to develop DAX measures.
19. Program to demonstrate Visualization types in Power BI.
20. Program to demonstrate and create a dashboard in Power BI.

Learning Resources:
1. https://learn.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started
2. Mastering Microsoft Power BI, Second Edition by Greg Deckler & Brett Powell.
3. Mastering Microsoft Power BI, Second Edition by Chandraish Sinha.

PYTHON FOR DATA SCIENCE LAB (LAB-II)

Credits: 2 **Semester: I**
Course Code: MDS22109 **No of Lecture Hours: 60**

Course Objective:

- Acquire the skill of Python Programing concepts

Course Outcome:

- Able to develop programs in Python.
- Able to implement functions using parameters.
- Understand the concept of expressions.
- Construct programs using Control structures
- List the operations that can be performed on strings.
- Identify the differences between lists and dictionaries.
- List the concepts of polymorphism and GUI programming

Lab Experiments:

1. Program to Demonstrate Data types and Operators
2. Program to Demonstrate Control structures
3. Program to Demonstrate Functions and recursive functions
4. Program to Demonstrate Lists and Dictionaries
5. Program to Demonstrate Tuples and sets
6. Program to Demonstrate Classes, Inheritance, and exceptional Handling
7. Program to Demonstrate packages NumPy, Pandas and Matplotlib
8. Program to Demonstrate Reading and handling missing data
9. Program to Demonstrate handling String operations
10. Program to Demonstrate File operations
11. Display Hello World after clicking a button
12. GUI program to display the message after selecting a radio button
13. GUI program to display a (i) Slider (ii) list box (iii) Scroll bar (iv) Check button
14. GUI program to display the message after entering a text into the entry box.

Note: Use of Python IDEs like PyCharm, Spyder or Anaconda should be used for executing programs.

MACHINE LEARNING LAB (LAB –III)

Credits: 2 **Semester: I**
Course Code: MDS26110 **No of Lecture Hours: 60**

Course Objective:
To acquire the skill of machine Learning techniques using Python

Course Outcome:
CO1: Learn to Load datasets.
CO2: Learn about the various libraries offered by Python to manipulate, preprocess, ars visualize data.
CO3: Learn the technique to reduce the number of variables using Feature Selection and Feature Extraction.
CO4: Learn in building models and model persistence using regression, classification.
CO5: Learn various machine learning algorithms like KNN, Decision Trees, SVM, and Clustering in detail.

Description:
- Data sets can be taken from standard repositories (https://archive.ics.uci.edu/ml/datasets.html) or Kaggle datasets or constructed by the students.

Lab Experiments:
1. Accessing different Datasets-Kaggle, UCI Machine Learning Repository, Google Dataset Search Engine, Microsoft datasets, and government datasets.
2. Data Preprocessing Implementation
3. Feature selection Implementation
4. Exploratory Data Analysis
5. Supervised Learning Algorithm Implementation-I
   i. Simple Linear Regression
   ii. Multiple Linear Regression
   iii. Polynomial Regression
   iv. Ridge Regression
   v. Lasso Regression
   vi. Supervised Learning Algorithm Implementation-II
   vii. Logistic Regression
   viii. Support Vector Machine Regression and Classification
   ix. K-Nearest Neighbor (KNN) – Regression and Classification
   x. Decision Tree Regression and Classification
   xi. Random Forest Regression and Classification
   xii. Naïve Bayes Classifier
6. Unsupervised Learning Algorithm Implementation k-means clustering

SECOND SEMESTER SCHEME

YEAR-WISE AND SEMESTER-WISE DISTRIBUTION OF SUBJECTS M.Sc. DATA SCIENCE
SECOND SEMESTER
ACADEMIC YEAR 2026-27 OF 2026-28 BATCH (CBCS)

Sl. No.	Subject Code	Title of the Subject	Hours /Week	Credits	Marks			Duration of Exam (hrs.)
					Internal	External	Total	
1	*MDS24201	* Human Values & Prof Ethics (**AEC-2**)	2	2*	20	30	50	2
2	*MDS26202	*Advanced Excel(**SEC-2**)	2	2*	50	-	50	-
3	MDS26203	Design and Analysis of Algorithms |**(GE-InterDisciplinary-1)**	4	4	40	60	100	3
DSC								
4	MDS22204A MDS22204B MDS24204C	Elective-I **(DSE-I)** Cloud Computing Social Media Analytics Web Analytics	4	4	40	60	100	3
5	MDS24205	Regression Analysis & Statistical Inference (**DSC6**)	4	4	40	60	100	3
6	MDS26206	Deep Learning (**DSC7**)	4	4	40	60	100	3
7	MDS26207	Advanced Databases (**DSC8**)	4	4	40	60	100	3
PRACTICALS								
7	MDS20208	Data Analytics Lab **(Lab-IV)**	4	2	40	60	100	3
8	MDS26209	Deep Learning Lab (**Lab-V**)	4	2	40	60	100	3
9	MDS26210	Advanced Databases Lab **(Lab-VI)**	4	2	40	60	100	3
Total			**36**	**26**	**390**	**510**	**900**	

***GE-Inter disciplinary (ID) paper will be offered by other Departments.**

*** Ability Enhancement Course (AEC) and Skill Enhancement courses (SEC) is excluded for SGPA and CGPA calculation, but it is compulsory to pass the examination.**

CIA components for Internal marks of theory paper: i) Assignment: 3M ii) Attendance: 4M iii) Group Discussion OR Presentation OR Case Study OR Computer Based Test: 5M iv) Mid Exam 1: 12.5M v) Mid Exam 2: 12.5M vi) Seminar or Viva Voce: 3M

CIA components for Practical Internal marks: i) Attendance: 5M ii) Observation/record book: 5M iii) Practical skill acquired: 10M iv) Pre final: 20M.

*Ability Enhancement Compulsory Course (AECC) * Discipline Specific Core (DSC)

HUMAN VALUES AND PROFESSIONAL ETHICS

Credits: 2 **Semester: II**
Course Code: MDS24201 **No of Lecture Hours: 30**

Course Objective:

- Develop a deep understanding of human and interpersonal values.
- Cultivate critical thinking skills to analyze situations, make informed decisions, and solve problems creatively.
- Analyze Ethical Theories and develop Ethical Decision-Making Skills.
- Develop organizational skills to maintain order and efficiency in both personal and professional life.
- Foster resilience to overcome challenges and setbacks, building a positive mindset in the face of adversity and promote a sense of global citizenship by understanding interconnectedness and global issues.

Course Outcome:

CO1: Students will be able to understand the concept of Universal Human Values.
CO2: Students will be able to analyze Ethical Theories and develop Ethical Decision-Making Skills
CO3: Students will be able to cultivate Interpersonal Skills and Communication.
CO4: Students will be able to examine Ethical Issues in Various Professions
CO5: Students will be able to promote Social Responsibility and reflect on Personal and Professional Values

Course Content	Hours Allotted	Pedagogy
Module I: Human Values • Positive affirmations, Assertiveness, Constructive criticisms • Human Relationship - formal and informal • Boundaries in relationships • Guilt and grudges • Forgiveness - accepting self & others.	**(6)**	Power point presentations / Lectures
Module II: Professional Values • Dealing with changes, adaptability skills • Transition from College to career • Work-life • Leader and manager • Integrity, Accountability, sustainability and Social responsibility	**(6)**	Case Studies / Review of research articles
Module III: Ethical Theories: • Distinction between ethics and morality • Common Ethical dilemmas in Profession • Ethical theories: Utilitarianism, Deontology, Virtue Ethics, Ethical Relativism, Consequentialism and Rights-based Ethics • Application of ethical theories in real-life dilemmas.	**(6)**	Assignments

Module IV: Ethical Decision Making: • Steps involved in Ethical decision-making. • Case-studies to analyze and resolve ethical dilemmas. • Problem solving skills - Process, approaches, and their components, • Lateral thinking, Tools and techniques, application of SMART analysis.	**(6)**	Lectures
Module V: Ethics in Professionalism: • Interpersonal skills for ethical collaboration in a diverse workplace. • Ethical issues and challenges in business, healthcare, technology, and academia. • Case studies to understand the ethical dimensions of professional practices. • Addressing biases and promoting inclusivity in professional interactions.	**(6)**	Assignments / Lectures

Learning Resources	
1.	Handout on Human Values and Professional Ethics - Loyola Academy

ADVANCED EXCEL

Credits: 2 **Semester: II**
Course Code: MDS26202 **No of Lecture Hours: 30**

Course Objective:

- To perform mathematical computations and data summarization using built-in mathematical functions.
- To understand and manage workbooks and worksheets effectively using Excel tools.
- To enhance data presentation and decision-making skills through advanced analytical tools.

Course Outcomes:

CO1: Develop accurate and efficient Excel-based solutions for real-world data analysis problems.
CO2: Design well-structured and error-free spreadsheets for academic and professional applications.
CO3: Develop visually effective and well-structured analytical reports using advanced Excel features.

Lab Practical

Functions and Advanced Tools

1. Implementing the Logical Functions (IF, AND, OR NOT).
2. Implementing the Text Functions (LEFT, RIGHT, MID, CONCATENATE).
3. Implementing Mathematical Functions (SUM, AVERAGE, ROUND).
4. Implementing the Statistical Functions (COUNT, MAX, MIN).
5. Implementing the Date Functions of (TODAY, NOW, DATE, TIME).
6. Implementing Information Functions (ISERROR, ISNUMBER, ISTEXT).
7. LOOK UP Functions VLookup, HLookup

Spreadsheet Management

8. Workbook and Worksheet Management (Open, Close, Save, Rename).
9. Cell Referencing (Absolute, Relative, Mixed).
10. Editing, Copying and Moving Cells.
11. To demonstrate Tables and Sorting using Excel.
12. To implement Defining and Managing Names in Excel.

Advanced Features

13. Conditional Formatting
14. Adding Filters
15. To create Pivot Tables and Pivot Charts for Data Summarization.
16. To create Combo Charts for Data Representation.
17. To create adding of Comments using Proofing Tools.
18. To create Page Layout options.
19. Macros

Learning Resources:

Textbook:

1. "Business Data Analysis using Excel", "David Whigham., Indian Edition. Oxford University Press. (Cost: Rs-7093/-)*

Reference books:

2. "Statistical Tools for Managers-using MS-Excel" Apte.D.P.2009. New Delhi: Excel Books. (Cost: Rs-621/-)*

DESIGN AND ANALYSIS OF ALGORITHMS

Credits: 4 **Semester: II**
Course Code: MDS26203 **No of Lecture Hours: 60**
Prerequisites: PYTHON PROGRAMMING
Course Objective:

- Analyze the asymptotic performance of algorithms.
- Write rigorous correctness proofs for algorithms.
- Demonstrate a familiarity with major algorithms and data structures.

Course Outcome:
CO1: Analyze worst-case running times of algorithms using asymptotic analysis.
CO2: Describe the divide-and-conquer paradigm and explain when an algorithmic design situation calls for it.
CO3: Describe the dynamic-programming paradigm and explain when an algorithmic design situation calls for it.
CO4: Describe the greedy paradigm and explain when an algorithmic design situation calls for it.
CO5: Understand NP-Hard and NP-Complete Problems

Course Content	Hours Allotted	Pedagogy
Module I: ● **Introduction to Algorithms**: Algorithm Specification, Performance Analysis ● **Elementary Data Structures:** Stacks and Queues, Trees, Dictionaries, Priority Queues, Sets and Disjoint Set Union, Graphs	**(12)**	Powerpoint presentations / Lectures
Module II: ● **Divide and Conquer:** Binary Search, Finding the Maximum and Minimum, Merge Sort; Quick Sort, Selection sort, Strassen's Matrix Multiplication, Convex Hull.	**(12)**	Case Studies / Review of research articles
Module III: ● **Greedy Method:** Knapsack Problem, Job Sequencing with Deadlines, Minimum-Cost Spanning Trees (Kruskal's & Prim's), Single Source Shortest Paths (Dijkstra's).	**(12)**	Assignments
Module IV: ● **Dynamic Programming:** General Method, All-Pairs Shortest Paths, SingleSource Shortest Paths, Optimal Binary Search Trees, 0/1 Knapsack, Traveling Salesperson Problem. **Back Tracking technique:** General Method, 8-Queens Problem, Graph Colouring, Hamiltonian Cycles	**(12)**	Lectures

Module V: **Branch-Bound technique:** General Method, 0/1 Knapsack Problem, Traveling Sales Person problem. **NP-Hard and NP-Complete Problems:** Basic Concepts, Cook's Theorem, NP-Hard and NPComplete problems. NP-Hard Scheduling Problems, Some Simplified NP-Hard Problems.	**(12)**	Assignments / Lectures

LearningResources	
1	E Horowitz, S Sahni, S Rajasekaran, (2007): Fundamentals of Computer Algorithms, 2/e, Universities Press.
2	T.H. Cormen, C.E. Leiserson, R.L Rivert, C Stein, (2010): Introduction to Algorithms, 3/e, PHI.
3	R. Pannerselvam (2007): Design and Analysis of Algorithms, PHI.
4	Hari Mohan Pandey, (2009): Design, Analysis and Algorithm, University Science Press.

CLOUD COMPUTING

Credits: 4 **Semester: II**
Course Code: MDS22204A **No of Lecture Hours: 60**
Prerequisites: Networking, Operating Systems

Course Objective:
- To introduce the fundamental information of cloud computing, what cloud computing is about, available services and key players, cloud computing benefits.

Course Outcome:
CO1: Illustrate the main concepts, features, challenges, and risks in cloud computing.
CO2: Describe virtualization of clusters and Data centers, virtual clusters, and resource management.
CO3: Identify the architectures over virtualized data centers.
CO4: Explain the core issues of cloud computing such as cloud security and trust management.
CO5: Compare various cloud programming and software environments.

Course Content	Hours Allotted	Pedagogy
Module I: Introduction to Cloud Computing • Cloud Computing in a Nutshell • System Models for Distributed and Cloud Computing • Roots of Cloud Computing • Grid and Cloud • Layers and Types of Clouds • Desired Features of a Cloud • Basic Principles of Cloud Computing, Challenges and Risks • Service Models	**(12)** 2 2 2 1 1 1 1 2	Power point presentations / Lectures
Module II: • Virtual Machines and Virtualization of Clusters and Data Centers • Levels of Virtualization, Virtualization Structures/Tools and Mechanisms • Virtualization of CPU, Memory, and I/O Devices • Virtual Clusters and Resource Management, Virtualization • Data-Centre Automation • Case studies: Xen Virtual machine monitors- Xen API • VMware - VMware products-VMware Features • Microsoft Virtual Server - Features of Microsoft Virtual Server	**(12)** 2 2 2 1 2 2 **1**	Case Studies / Review of research articles
Module III: Cloud computing architectures over Virtualized Data Centers: • Data-Centre design and Interconnection networks • Architectural Design of Compute and Storage Clouds • Public Cloud Platforms, GAE, AWS, Azure • Inter-cloud Resource Management	**(6)** 3 2 2 5	Assignments

Module IV: Cloud Security and Trust Management, Data Security in the Cloud: • An Introduction to the Idea of Data Security • The Current State of Data Security in the Cloud • Homo Sapiens and Digital Information • Cloud Computing and Data Security Risk • Cloud Computing and Identity • The Cloud, Digital Identity, and Data Security • Content Level Security-Pros and Cons	**(12)** 3 1 1 1 2 2 **2**	Lectures
Module V: Cloud Programming and Software Environments: • Features of Cloud and Grid Platforms • Parallel and distributed Programming Paradigms • Programming Support of Google App Engine • Programming on Amazon AWS and Microsoft Azure • Emerging Cloud Software Environments	**(12)** 2 3 2 3 2	Assignments / Lectures

Learning Resources	
1.	RajkumarBuyya, James Broberg and Andrzej M. Goscinski..Cloud Computing: Principles and Paradigms (Wiley Series on Parallel and Distributed Computing). Wiley Publishing.2011(unit I, IV some part)
2.	Kai Hwang. Geoffrey C.Fox, Jack J. Dongarra, Distributed and Cloud Computing From Parallel Processing to the Internet of Things.2012. Elsevier. (unit-II, unit-III, unit-V)
3.	Raluca Ada Popa, Catherine M.S. Redfield, NickolaiZeldovich, and HariBalakrishnan. CryptDB: Protecting Confidentiality with encrypted Query Processing. 23rd ACM
4.	Symposium on Operating Systems Principles (SOSP 2011), Cascais, Portugal. Craig Gentry. September A Fully Homomorhic Encryption Scheme.2009.

SOCIAL MEDIA ANALYTICS

Credits: 4 **Semester: II**
Course Code: MDS22204B **No of Lecture Hours: 60**

Course Objective:

- The objective of this course is to enhance the skills of the student in social media, to extract data from social media platforms and transform them into actions resulting in improved business values.

Course Outcome:

CO1: Identify various platforms in social media.
CO2: Understand processing of social media.
CO3: Compare differences between twitter and other social media networks.
CO4: Analyze Facebook information and write business cases.
CO5: Differentiate social media networks Instagram (i.e., usage of Instagram and data processing techniques also they will get idea)

Course Content	Hours Allotted	Pedagogy
Module I: • Social Media and Its Importance • Various Social Media Platforms • Social Media Mining • Challenges for Social Media Mining • Social Media Mining Techniques	**(12)** 2 2 2 3 3	Power point presentations / Lectures
Module II: • Types of Social media Networks • Generic process of social media mining • Preprocessing and cleaning in R • Data Modeling Steps • Sentiment Analysis • Steps in sentiment analysis • Business Case	**(12)** 2 2 2 2 2 2	Case Studies / Review of research articles
Module III: **Twitter:** • Twitter and its importance • Understanding Twitters APIs • Collecting Twitter user features from APIs • Twitter vocabulary • Twitter sentiment analysis: Business Case	**(12)** 3 3 3 3	Assignments
Module IV: **Facebook:** • Social network analysis and visualization • Getting Facebook page data • Trend analysis • Spam detection • Business Case **Instagram:**	**(12)** 1 1 2 1 1	Lectures

• Access Data from Building a dataset • Popular personalities • Finding the most popular destination • Clustering the pictures • Recommendations to the users, Business case	1 1 1 1 1	
Module V: More Social Media Websites: • Searching on Social Media, Accessing Product Reviews from Sites • Retrieving Data from Wikipedia, Accessing Data from Quora • Mapping Solutions from Google Maps • Professional Network Data from LinkedIn • Getting Blogger data	**(12)** 3 3 2 2 2	Assignments / Lectures

Learning Resources	
1.	Ravindran, Sharan Kumar. Garg, Vikram. 2015. Mastering Social Media Mining with R.
2.	Social Media Data Mining and Analytics, Gabor Szabo, Gungor Polatkan, P. Oscar Boykin 2018 Wiley publishers

WEB ANALYTICS

Credits: 4 **Semester: II**
Course Code: MDS24204C **No of Lecture Hours: 60**

Course Objective:

- To learn web analytics from a strategic and practical perspective.
- Explore different types of analytics and why they are important for business.
- Learn various web analytics processes and metrics used to measure.
- Learn techniques using Google Web analytics, traffic analysis, click path analysis and segmentation.

Course Outcome:

CO1: Learn how to use web analytics to answer the complicated questions facing UX professionals.
CO2: Able to measure a website's effectiveness.
CO3: Learn methods about how users navigate a website.

Course Content	Hours Allotted	Pedagogy
Module I: • **Web Analytics Present and Future:** A brief history of web analytics, Current Landscape and Challenges, Traditional Web analytics, Web analytics Activities, Measuring, Trinity. • **Data collection:** Understanding the Data Landscape, Click stream Data, Outcomes Data, Research Data, Competitive Data.	**(12)** 8 4	Power point presentations / Lectures
Module II: • **Introduction to Web Analytics:** Definition, User Experience and Web Analytics Questions. • **Web Analytics Approach:** Introduction, A model of Analysis, Showcasing the work, Context Matters, Contradicting the data. • **Working of Web Analytics:** Introduction, Log File Analysis, Page Tagging, Metrics and Dimensions, Interacting with data in Google Analytics.	**(12)** 4 4 4	Case Studies / Review of research articles
Module III: • **Goals:** Introduction, Definition of Goals and Conversions, Conversion Rate, Goal Reports in Google Analytics, Finding the right things to measure as key, Performance Indicators, Measure on a website that can constitute a goal. • **Learning about users:** Introduction, Visitor Analysis.	**(12)** 8 4	Assignments
Module IV: Title: • **Traffic Analysis:** Introduction, Source and medium, Organic Search, Search Query Analysis, Referral Traffic, Direct Traffic, Paid Search Keyword. • **Analyzing usage of content:** Introduction, Website content Reports.	**(12)** 8 4	Lectures
Module V:	**(12)**	Assignments /

• **Click-Path Analysis:** Introduction, Focus on Relationships between pages, Navigation Summary, Visitors Flow Report, Analyzing how users move from one page type to another. • **Segmentation:** Introduction, Necessity, Procedure to segment, • Ways to Segment, Useful ways to segment UX questions.	4 4 **4**	Lectures

Learning Resources	
1.	Web Analytics an hour a Day by Avinash Kaushik, Sybex, 1st Edition, 2007. (Unit – I)
2.	Practical Web Analytics for User Experience by Michael Beasley, Morgan Kaufmann, 1st Edition, 2013. (Unit II to Unit V)
3.	Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity by Avinash Kaushik, 1st edition, Sybex, 2009.
4.	Web Data Mining: Exploring Hyperlinks, Content and Usage Data by Bing Liu, 2nd Edition, Springer, 2011.
5.	Google Analytics by Justin Cutroni, O'Reilly, 2010.

REGRESSION ANALYSIS & STATISTICAL INFERENCE

Credits: 4 **Semester: II**
Course Code: MDS24205 **No of Lecture Hours: 60**
Prerequisite: Probability Theory and Distribution

Course Objective:

- Use data to make evidence-based decisions and communicate the purpose of analysis. findings from the analysis and the implications of those findings.
- Develop an understanding developing the relationships between variables and investigate possible diagnostics, formulate the feasible solution using a regression model in real – life problems.

Course Outcome:

CO1: Understand the basic concepts of hypothesis and use large - sample tests.
CO2: Use small samples and non - parametric tests of hypothesis to infer the data.
CO3: Develop an in depth understanding of linear regression models.
CO4: Predict the parameters and fit the model, investigate possible diagnostics in regression modelling and analysis.
CO5: Understand generalization of linear models and demonstrate model selection techniques.

Course Content	Hours Allotted	Pedagogy
Module I: **Hypothesis Testing:** • Basic concepts - population, sample, parameter, statistic. Hours • Sampling, types of sampling, sampling distribution, standard error, test of significance. • Hypothesis - types of hypotheses, errors in hypothesis, critical region, level of significance, power of the test - significance test using p- value and table value. • **Large Sample Tests:** Concept of large and small samples. Large sample test for single and difference of means • (for known and unknown values of σ) • Large sample test for single and difference of variance. • Large sample test for single and difference of proportions.	**(12)** 1 2 2 2 1 2 2 2	Power point presentations / Lectures
Module II: **Small Sample Tests (Description and applications only - No derivations):** • t - test for single mean, paired means difference of means (independent samples). • t- test for single proportion. • F - test for equality of two variances. • ANOVA one way and two-way classification without interactions. • Chi - square test for independence of attributes and goodness of fit and single variance. **Non-Parametric Tests (Description, properties, and applications only - No derivations):**	**(12)** 2 1 1 1 2	Case Studies / Review of research articles

• Concept of non - parametric tests, assumptions, and limitations of non - parametric tests. • Run test for randomness for single and two samples. • Sign test and Wilcoxon signed rank test for single and paired samples. • Median tests and Mann - Whitney - Wilcoxon tests for two samples. • One sample Kolmogorov Smirnov test, Kruskal- Wallis H test	1 1 1 1 1	
Module III: Linear Regression Analysis **Simple Linear Regression** • Introduction to a linear and nonlinear model. Ordinary Least Square methods. • Simple linear regression model to describe a linear relationship. • Fitting a linear trend to time series data, • Validating a simple regression model using t, F, and p test. • Developing confidence interval. Precautions in interpreting regression results. **Multiple Linear Regression:** • Concept of Multiple regression model to describe a linear relationship. • Assessing the fit of the regression line, inferences from multiple regression analysis. • Problem of overfitting a model, comparing two regression models, • Prediction with multiple regression equations. • Stepwise regression analysis - backward and forward analysis. • Information criteria: AIC and SBC. • Criteria for choice of subset size: coefficient of determination, adjusted coefficient of determination, residual mean square, Mallow's Cp, statistic, model validation.	**(12)** 3 3 2 1 3	Assignments
Module IV: Performance Testing: **Model Adequacy:** • Residual analysis, PRESS statistics. • Detection, and treatment of outliers. • Lack of fit of the regression model, test of lack of fit. • Problem of autocorrelation and heteroscedasticity. • Estimation of pure errors from near neighbours. **Multicollinearity:** • Introduction, sources of multicollinearity, effects of multicollinearity. **Multicollinearity diagnostics: examination of correlation matrix.** • Variance Inflation factors (VIF), • Eigen system analysis of X'X. **Methods of dealing with Multicollinearity:** • Collecting additional data. • Model re-specification, and ridge regression.	**(12)** 5 3 2 2	Lectures
Module V: Other Regression Models:	**(12)**	Assignments /

Generalized linear model: link functions and linear predictors. • Parameter estimation and inference in the GLM • Prediction and estimation with the GLM • Residual Analysis, and concept of over dispersion. **Model Building and Non - Linear Regression Models:** Variable selection • Model building, model misspecification. **Model validation techniques: Analysis of model coefficients** • Predicted values, data splitting method. • Nonlinear regression model, nonlinear least squares. • Transformation to linear model • Parameter estimation in a non-linear system. • Statistical inference in nonlinear regression.	4 2 4	Lectures

Learning Resources	
1.	Manoj Kumar Srivastava and Namita Srivastava, Statistical Inference - Testing of Hypotheses. Prentice Hall of India, 2014.
2.	Robert V Hogg. Elliot A Tannis and Dale L. Zimmerman, Probability and Statistical Inference, 9th edition, Pearson publishers, 2013.
3.	Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, Introduction to Linear Regression Analysis, Third Ed., Wiley India Pvt. Ltd., 2016.
4.	Norman R. Draper. Harry Smith: Applied Regression Analysis, WILEY India Pvt. Ltd. New Delhi: Third Edition, 2015.

Suggested Readings:

1. Mare S. Paolella, Fundamental statistical inference: A computational approach, Wiley, 2018.
2. B. K. Kale and K. Muralidharan, Parametric Inference, Narosa Publishing House, 2016.
3. Miller, I and Miller, M, John E. Freund's Mathematical statistics with Applications, Pearson Education, 2002.
4. Rao, C.R., Linear Statistical Inference and its applications, 2nd Edition, Wiley Eastern, 1973.
5. Gibbons. J.D., Non-Parametric Statistical Inference, 2/e,Marckel Decker, 1985.
6. Bansilal, Sanjay Arora and Sudha Arora, Introducing Probability and Statistics, 2/e, Satya Prakash Publications, 2006.
7. George Casella and Roger L.Berger:, Statistical Inference, 2nd edition, Casebound Engelska, 2002.
8. Johnson, R A., Wichern, D. W., Applied Multivariate Statistical Analysis, Sixth Ed., PHI learning Pvt., Ltd., 2013.
9. Jain Pardoe, Applied Regression Modeling, John Wiley and Sons, Inc, 2012.

DEEP LEARNING

Credits: 4 **Semester: II**
Course Code: MDS26206 **No of Lecture Hours: 60**
Prerequisites: Machine Learning

Course Objective:

To build the foundation of neural networks.
Explore advanced topics of neural networks.
To design and analyze various algorithms and techniques with a modern outlook.
To enable the students, develop successful deep learning projects.

Course Outcome:

CO1: Learn the fundamental principles of deep learning.
CO2: Identify the deep learning algorithms for various types of learning tasks in various domains.
CO3: To explore Deep learning techniques and various feature extraction strategies.
CO4: To mathematically understand the deep learning approaches and paradigms.
CO5: Implement deep learning algorithms and solve real-world problems.

Course Content	Hours Allotted	Pedagogy
Module I: Introduction:	**(12)**	Power point presentations / Lectures
• Deep Learning: Introduction, Historical trends and Challenges motivating Deep Learning	2	
• Supervised Learning Algorithms, Unsupervised Learning Algorithms	2	
• Linear Algebra: Scalars, Vectors, Matrices and Tensors	2	
• Gradient Based Optimization, Optimizing Deep Learning Models: Batch and Mini-batch Algorithms	2	
• Challenges in Neural Network Optimization: local minima, Plateaus and Saddle Points	2	
• Stochastic gradient descent, Momentum, and Nesterov Momentum method	2	
Module II: Neural Nerworks:	**(12)**	Case Studies / Review of research articles
• Feedforward neural networks	2	
• Deep Networks	2	
• Architecture Design	2	
• Regularizing a Deep network	2	
• Model exploration and hyperparameter tuning.	4	
Module III: Convolution Neural Networks::	**(12)**	Assignments
• Introduction to convolution neural networks	2	
• Convolution Operation	2	
• Motivation, Pooling	2	
• Convolution and Pooling as an Infinitely strong prior	2	
• Variants of the basic Convolution function	2	

• Structured outputs, Efficient Convolution Algorithms	2	
Module IV:Sequence Modeling : Recurrent Nets and Recursive Nets: • Unfolding computational graphs, recurrent neural networks (RNNs) • Bidirectional RNNs • Encoder-Decoder sequence to sequence architectures • Deep Recurrent networks • Recursive Networks • LSTM- The Long Short-Term Memory	**(12)** 2 2 2 2 2 2	Lectures
Module V: Auto encoders: • Under complete auto encoders • Regularized auto encoders, sparse auto encoders, denoising auto encoders • Representational power, layer, size, and depth of auto encoders • Stochastic encoders and decoders • Application of auto encoders	**(12)** 2 2 2 2 4	Assignments / Lectures

Learning Resources	
1.	Deep Learning by Ian Goodfellow, YoshuaBengio. Aaron Courville. The MIT Press, 2016
2.	Deep Learning, A Practical Approach by Rajiv Chopra, Khanna Book Publishing, 2018
3.	Jeff Heaton, Deep Learning and Neural Networks, Heaton Research Inc, 2015

ADVANCED DATABASES

Credits: 4 **Semester: II**
Course Code: MDS26207 **No of Lecture Hours: 60**
Prerequisites: Basic Database concepts

Course Objective:

- Relate the role of a database management system in an organization.
- Demonstrate basic database concepts, including the structure and operation of the relational data model, non-relational data model.
- Explain and successfully apply logical database design principles, including E-R diagrams and database normalization.
- Demonstrate the concept of a database transaction, hashing, indexing.
- Explain different types of database architecture.
- Explain the concept of parallel database & distributed databases.

Course Outcome:

CO1: Design a data base for a system using E-R data model and Relational Data model.
CO2: Design logical database with all integrity constraints over relations.
CO3: Apply normalization steps in database design and removal of data anomalies.
CO4: Extend the characteristics of database transactions.
CO5: Distinguish the different types of NoSQL databases

Course Content	Hours Allotted	Pedagogy
Module I: ● Introduction to Relational Database Model: Structure of Relational Database, Database Schema, Keys, Schema Diagrams, Relational Query Languages, Relational Operations. ● Database design and the E-R Model: Overview of the Design Process, the Entity-Relationship model, Constraints, Removing Redundant Attributes in Entity sets, Entity-Relationship Diagrams, Reduction to relational schemas, Entity-Relationship design issues, Extended E-R features - Specialization and generalization.	**(12)** 6 6	Powerpoint presentations / Lectures
Module II: ● Relational Database Design: Features of Good Relational Designs, Atomic Domains and First Normal Form, Decomposition using Functional Dependencies, Functional-Dependency Theory, Decomposition using Multi-valued Dependencies, More Normal Forms, Database-Design Process and Modeling Temporal Data ● Indexing and Hashing: Basic concepts, Ordered Indices, B+ tree index files, B+ tree extensions, Multiple-key access, Static Hashing, Dynamic Hashing, Comparison of ordered Indexing and	**(12)** 6 6	Case Studies / Review of research articles

Hashing, Bitmap indices, Index definition.		
Module III: ● Transactions: Transaction Concept, a Simple Transaction Model, Storage Structure, Transaction Atomicity and Durability, Transaction Isolation, Serializability, Transaction Isolation Levels, Implementation of Isolation Levels, ● Database System Architecture: Centralized and Client –Server Architectures, Server System Architectures, Parallel Systems, Distributed Systems, Network Types.	**(12)** 8 4	Assignments
Module IV: ● Parallel Database: Introduction, I/O Parallelism, Inter query Parallelism, Intra query Parallelism, Intra Operation Parallelism, Interoperation Parallelism, Query Optimization, Design of Parallel Systems, Parallelism on Multi Core Processor. ● Distributed Databases: Homogeneous and Heterogeneous Databases, Distributed Data Storage, Distributed Transactions, Commit Protocols. ● NOSQL: Value of Relational Databases, impedance mismatch, Application, and Integration Databases, Attack of the clusters, Emergence of NoSQL. ● Aggregate Data Models: Aggregates, Key-Value and Document Data Models, Column Family Stores.	**(12)** 4 3 3 2	Lectures
Module V: ● More Details on Data Models: Relationships, Graph Databases, Schema less Databases, Materialized Views, Modeling for Data Access. ● Distributed Models: Single Server, Sharding, Master-Slave Replication, and Peer to Peer Replication. ● Consistency: Update Consistency, Read Consistency, Relaxing Consistency, Relaxing Durability. ● Key-Value Databases: Introduction to Key Value Store, Features, Structure of Data, Scaling, Use-case: Storing webpages in a key-value store ● Document Database: Introduction, Features. ● Column -Family (Big-table) Databases Column family basics, Understanding Column family keys, Benefits of Column family Systems, Case study: Storing analytical information in a Bigtable, Google maps ● Graph Databases: Introduction, Features.	**(12)** 2 2 2 2 2 2	Assignments / Lectures

Learning Resources	
1.	DatabaseSystemConceptsbyAbrahamSilberschatz,HenryF.korth,S.Sudarshan, McGrawHill, Sixth Edition, 2011. (Units I –III, IV).
2.	NoSQLDistilledbyPramodJSadalage,MartinFowler,Addison-WesleyProfessional;1st edition, 2012. (Unit – IV, V)
3	Making Sense of NoSQL: A guide for managers and rest of us; Dan McCreary and Ann Kelly, 2014
4.	FundamentalsofDatabaseSystemsbyRamezElmasri,ShamkantB.Navathe,Addison-Wesley, Sixth Edition, 2011.

DATA ANALYTICS LAB (LAB IV)

Credits: 2 **Semester: II**
Course Code: MDS20208 **No of Lecture Hours: 60**

Course Objective:

- Data Analytics Lab is designed to introduce the concepts of diagrammatic representation of data, descriptive statistics, correlation and regression analysis, and theory of estimation.
- Data Analytics Lab also deals with the concept of parametric tests for small samples.

Course Outcome:

- Apply the concepts of graphical representation, descriptive statistics, correlation, and regression analysis tools for the statistical data.
- Demonstrate the concepts of point and interval estimation of unknown parameters and their significance using small samples.
- Apply the idea of sampling distributions of difference statistics in testing of hypotheses.

List of Practicals:

1. Graphical representation of data using Histogram.
2. Graphical representation of data using Frequency curve.
3. Graphical representation of data using Ogives.
4. Computation of central tendencies (mean, median, mode, geometric mean and harmonic mean).
5. Computation of partition value - quartiles.
6. Computation of dispersion (mean deviation, quartile deviation, standard deviation and coefficient of variance).
7. Computation of coefficient of skewness.
8. Computation of correlation coefficient for two variables.
9. Computation of multiple correlation coefficient.
10. Computation of regression lines x on y and y on x for two variables.
11. Computation of multiple regression lines and regression coefficients.
12. Fitting of second-degree parabola curve.
13. Fitting of power curve.
14. Fitting of the exponential curve.
15. Estimation of residuals.
16. Estimation of R2 and R^2 for model selection.
17. Determining Logistic regression
18. Determining Poisson Regression.
19. Generating random numbers for discrete series (Binomial and Poisson).
20. Generation of random numbers for continuous series (Uniform, Exponential and Normal).
21. Maximum likelihood estimation for discrete and continuous distributions.
22. t- test for single mean, difference of means (for independent samples).
23. Paired t - test for difference of means.
24. F- test for testing the significance of difference of two population variances.

25. ANOVA for one - way classification of data (both equal and unequal number of observations).
26. ANOVA for two - way classification of data.
27. Chi – square test for testing goodness of fit.
28. Chi – square test for testing independence of attributes.

Learning Resources	
1.	Hands-on programming with R, Garrett Grolemund, O'Reilley, 1st Edition, 2014.
2.	R for everyone, Jared Lander, Pearson, 1st Edition, 2014.
3.	Statistics: An Introduction Using R, Michael J. Crawley, WILEY, Second Edition, 2015.

DEEP LEARNING LAB (LAB V)

Credits: 2 **Semester: II**
Course Code: MDS26209 **No of Lecture Hours: 60**

Course Objective:

To Build The Foundation Of Deep Learning.
To Understand How To Build The Neural Network.
To enable students to develop successful machine learning concepts.

Course Outcome:

CO1: Implement deep neural networks to solve real world problems.
CO2: Choose appropriate pre-trained model to solve real time problem.
CO3: Interpret the results of two different deep learning models.

Lab Exercises:

1. Implement multilayer perceptron algorithm for MNIST Hand written Digit Classification.
2. Design a neural network for classifying movie reviews (Binary Classification) using IMDB dataset.
3. Design a neural Network for classifying news wires (Multi class classification) using Reuters dataset.
4. Design a neural network for predicting house prices using Boston Housing Price dataset.
5. Build a Convolution Neural Network for MNIST Hand written Digit Classification.
6. Build a Convolution Neural Network for simple image (dogs and Cats) Classification
7. Use a pre-trained convolution neural network (VGG16) for image classification.
8. Implement one hot encoding of words or characters.
9. Implement word embedding for IMDB dataset.
10. Implement a Recurrent Neural Network for IMDB movie review classification problem.

TEXT BOOKS:

• Reza Zadeh and BharathRamsundar, “Tensorflow for Deep Learning”, O‟Reilly publishers, 2018

REFERENCES:

• https://github.com/fchollet/deep-learning-with-python-notebooks

ADVANCED DATABASES LAB (LAB-VI)

Credits: 2 **Semester: II**

Course Code: MDS26210 **No of Lecture Hours: 60**

Course Objective:

- Acquire skill of database schema design, SQL Commands, NoSQL operations of MongoDB.

Course Outcome:

- Design and implement a database schema for a given problem-domain.
- Populate and query a database using SQL DML/DDL commands.
- Declare and enforce integrity constraints on a database
- Retrieve data using different SQL joins, subqueries, and correlated queries.
- Utilize the techniques used to create, insert, update, and delete data/documents.
- Utilize various techniques used to query the database.
- Utilize techniques to optimize querying using indexing.
- Apply methods to analyze data using aggregation techniques.
- Adopt knowledge about the role of NoSQL in business.
- Identify technique of splitting data across machines via sharding.

Lab Experiments:

a) Relational Databases:
   1. Implement DDL Statements.
   2. Implement DML Statements.
   3. Write the queries for implementing Built-In Functions.
   4. Design a Database and create required tables. Apply the constraints like Primary Key, Foreign Key, Not Null, and other constraints to the tables. Perform SQL Queries.
   5. Write the queries to implement the Joins.
   6. Write the queries to implement subqueries and correlated queries.

b) NoSQL Databases:
   1. Demonstrational operators in mongo dB to query the document.
   2. Demonstrate the methods to analyze data using aggregation techniques.
   3. Explain the techniques of splitting data across machines.
   4. Demonstrate LIMIT () and SKIP () methods in mongo db.
   5. Demonstrate example for Text Searching in Mongo db.
   6. Demonstrate the usage different remove methodologies to remove documents from collections.

• Select various design aspects and operations of MongoDB.
• Define objects, load data, query data and performance tune Key-Value Pair NoSQL databases.

Learning Resources	
1.	SQL, PL/SQL the Programming Language of Oracle by Ivan Bayross, BPB Publications, 4th Edition, 2010.
2.	MongoDB: The Definitive Guide by Kristina Chodorow, Shroff publisher, 2nd edition, 2013.
3.	NoSQL with Mongo DB in 24 hours by Brad Dayley, Pearson Education, 1st edition, 2015.

THIRD SEMESTER SCHEME

YEAR-WISE AND SEMESTER-WISE DISTRIBUTION OF SUBJECTS M.Sc. DATA SCIENCE
THIRD SEMESTER
ACADEMIC YEAR 2025-26 OF 2024-26 BATCH (CBCS)

Sl. No.	Subject Code	Title of the Subject	Hours /Week	Duration of Exam (hrs.)	Marks			Credits
					Internal	External	Total	
1	*MDS21301	*Soft skills **(AECC-3)**	2	-	50	-	50	2*
2	#MDS21302	Mini Project **(SEC-3)**	2	-	25	50	75	2
CORE								
3	MDS25303	Time Series & Forecasting Techniques **(Core-9)**	4	3	40	60	100	4
4	MDS23304 MDS23304A MDS23304B MDS23304C	Elective-II **(DSE-II)** Big Data Security Reinforcement Learning Scalable Architectures for ML Applications	4	3	40	60	100	4
5	MDS25305	Internet of Things **(Core-10)**	4	3	40	60	100	4
6	MDS25306	Deep Learning **(Core-11)**	4	3	40	60	100	4
7	MDS25307	Natural Language Processing **(Core-12)**	4	3	40	60	100	4
PRACTICALS								
7	MDS21308	Internet of Things Lab **(Lab-VII)**	4	3	40	60	100	2
8	MDS23309	Deep Learning Lab **(Lab-VIII)**	4	3	40	60	100	2
9	MDS25310	Natural Language Processing Lab **(Lab-IX)**	4	3	40	60	100	2
Total			**36**		**395**	**530**	**925**	**28**

*** Ability Enhancement Compulsory Course (AECC) is excluded for SGPA and CGPA calculation, but it is compulsory to pass the examination.**
#In the third Semester students will take one month summer project of 2 credits.

CIA components for Internal marks of theory paper: i) Assignment: 3M ii) Attendance: 4M iii) Group Discussion OR Presentation OR Case Study OR Computer Based Test: 5M iv) Mid Exam 1: 12.5M v) Mid Exam 2: 12.5M vi) Seminar or Viva Voce: 3M
CIA components for Practical Internal marks: i) Attendance: 5M ii) Observation/record book: 5M iii) Practical skill acquired: 10M iv) Pre final: 20M.
Remedial class/Communication skills/ Mentoring will be given 3 extra classes in a week.

*Ability Enhancement Compulsory Course (AECC) * Discipline Specific Core (DSC)

SOFT SKILLS

Credits: 2 **Semester: III**
Course Code: MDS21301 **No of Lecture Hours: 30**

Course Objective:

- To equip the student with the adequate soft skills required for any organization. And skills required enhancing his/her career.

Course Outcome:

- The student will be able to develop effective communication skills, presentation skills, interpersonal skills, team management skills, and leadership skills.

Course Content	Hours Allotted	Pedagogy
Module I: Goal Setting • Meaning of goal and goal setting short. • Medium- and long-term goal setting • Importance of goal setting-Choices/Selection of setting goals-Steps for goal setting-SMART goals	**(6)** 2 2 2	Power point presentations / Lectures
Module II: Time Management • What and why of time management • Necessity and benefits of time management-tools of time management • How to manage time wisely	**(6)** 2 2 2	Case Studies
Module III: Etiquettes • Get the first impression well. • Greet others & introduce yourself. • Body language- speak well. • Dressing sense- appeals to others.	**(6)** 2 2 2 2	Case Studies
Module IV: Group Discussion Skills: • Leadership Skills, Interpersonal Skills. • Persuasive Skills, Problem Solving Skills. • Conceptualization Skills. • Initiating the discussion, listening to others point of view.	**(6)** 2 2 2 2	Case Studies/ Lectures
Module V: Interview Skills: • Creating first impression in an interview, walkup to interview room. • How to approach the interview members, sitting posture in the interview room, body language	**(6)** 2 4	Assignments / Lectures

Learning Resources	
1.	Atkinson Frank.2011. Get More Out of Your Day.India:Vikas Books Private Limited. (Unit I in Pg 17-29, Unit II in Pg 5-16 and Pg 73 – 118)
2.	Butterfield Jeff. Soft Skills for Everyone. India:Cengage Learning. (Unit III in Pg 177-204 and Pg 336 -339, Unit IV in Pg 357 -373)

3.	Philips R. Hunsaker. Training in Interpersonal Skills. New Delhi:McGraw Hill. (Unit V in Pg 176 - 214)
4.	Time Management. USA:Harvard Business School Press. Boston.
5.	Padhy Kishore C and madhuchhanda. 2008.A to Z of Interview. India:Himalaya Publishing House.

MINI PROJECT

Credits: 2 **Semester: III**
Course Code: MDS21302 **No of Lecture Hours: 30**

Course Objective:

- The objective of mini project work is to develop quality analytical solutions. The primary emphasis of the project work is to understand and gain the knowledge of the principles of data science and analytics practices.

Course Outcome:

- The program prepares the students to take up positions as Systems Analysts, Systems Designers, Data scientists, Programmers and Project Managers in any field related to data science and analytics.

- The main aim of this mini project is to check the learning skills of students in Python, Machine Learning and Data Mining etc., which is covered in the I and II semesters. The students are expected to do a Mini project in the summer vacation following the2ndsemester.Project must be carried out by each student individually in a period of 30 days of duration. Students should submit a synopsis at the end of 2nd semester in consultation with the Project Guide. The synopsis should consist of the definition of the problem, scope of the problem and plan of action.

- Students are supposed to take this project work very seriously, as these efforts will be the showcase of their skill and knowledge in Python, Machine Learning and Data Mining etc. The mini project should be genuine and original in nature and should not be copied from anywhere else. If found copied, then it will be canceled automatically.

- After completion of project, in the 3rd semester students are required to present a Project Seminar covering the aspects of Data Collection, Data Visualization, Labeling, Data Selection, Data Pre-processing, Data transformation, Model Training, Model evaluation, Model testing and accuracy are required to show in the project work.

- A committee consisting of two or three faculty members of the department along with a guide will evaluate the project and award internal marks.

TIME SERIES ANALYSIS AND FORECASTING TECHNIQUES

Credits: 4 **Semester: III**
Course Code: MDS25303 **No of Lecture Hours: 60**

Course Objective:
- To understand the differences between cross-sections and time series and specific economic problems which occur while working with data of these types.

Course Outcome:
CO1: Identifying linear, quadratic, Gompertz and Logistic models where appropriate and describe models for seasonal variation. Also explains the methods used to study cyclic components.
CO2: Estimate seasonal effects of time-series data by using Winten's method, Brown's, Box – Jenkin's three-parameter exponential smoothing method**.**
CO3: To utilize AR, ARIMA models for time series data and to forecast the data using these models.
CO4: To interpret SARIMA model and criterion used to study them.
CO5: To explain and verify mathematical considerations for analysing time series, including concepts of stationarity, autocovariance, autocorrelation.

Course Content	Hours Allotted	Pedagogy
Module I:	**(12)**	Power point presentations / Lectures
Growth Models:		
• Modified Exponential curve, Gompertz curve, Logistic curve, and their fitting.	5	
Measurement of cyclic Variations:		
• Harmonic analysis, Auto – regression series – Markoff and Yule's series.	5	
• Periodogram and correlogram analysis.	2	
Module II:	**(12)**	Case Studies / Review of research articles
Measurement of irregular Component:		
• Variant difference method.	2	
Exponential Smoothing Methods:		
• Trend adjusted exponential smoothing. Double and triple exponential smoothing.	5	
• Winter's method. Brown's one parameter adaptive method.	1	
• Box – Jenkin's three parameter smoothing.	2	
• Harrison's Harmonic smoothing methods and Tracking signal.	2	
Module III:	**(12)**	Assignments
• Stationary time series. Autocovariance and autocorrelation functions.	2	
• Regression models for general time series data.	2	
• Detecting autocorrelation – Durbin – Watson test.	2	
• Estimating the parameters in time series regression models.	2	
• ARCH and GARCH modeling of non – linear time series	4	

data.		
Module IV: • Box – Jenkin's Time Series Methods: Moving average, Autoregressive (AR). • ARMA and AR integrated MA (ARIMA) models. • Estimation of ARIMA model parameters forecasting with ARIMA models.	**(12)** 4 4 4	Lectures
Module V: Title: • SARIMA (Seasonal Autoregressive Integrated Moving Average method). • Akaike's Information Criterion (AIC) • Bayesian Information Criterion (BIC) • Root Mean Square Error (RMSE) • Diagnostic checking of models. • Analysis of residuals.	**(12)** 2 2 2 2 2 2	Assignments / Lectures

Learning Resources	
1.	Box, G.E.P and Jenkins, GM: Time Series Analysis Forecasting and Control, Holden Day, San Francisco.
2.	Anderson, T.W.: The Statistical Analysis of Time series, John wiley, New York.
3.	Thomopouls, N.T.: Applied Forecasting methods. EngleWood cliffs, N.J, Prentice Hall.
4.	Wheel Wishart S.C. and S. Makridaks: Forecasting Methods for Management, 3rd Edition, John Wiley, New York.
5.	Sullivan, William G. and Wayne Claycambe. W: Fundamentals of Forecasting, Prentice Hall.

BIG DATA SECURITY

Credits: 4 **Semester: III**
Course Code: MDS23304A **No of Lecture Hours: 60**
Prerequisites: Big Data Analytics

Course Objective:
- Introduction to big privacy, ethics and security
- Learn about steps to secure big data and challenges
- Understand Hadoop Kerberos security implementation & configuration.
- Learn how to secure different Hadoop ecosystem components

Course Outcome:
CO1: Learn the fundamentals of big data security.
CO2: To explore the steps to secure big data.
CO3: Understand the different Hadoop security designs.
CO4: Implementing security for Hadoop ecosystem components.
CO5: Understand integrating Hadoop with Enterprise Security Systems and securing sensitive data in Hadoop.

Course Content	Hours Allotted	Pedagogy
Module I: Big Data Privacy, Ethics and Security : • Privacy, Re Identification of Anonymous People, • Why is Big Data Privacy self-regulating? • Ethics, Ownership, Ethical Guidelines • Big Data Security, Organizational Security.	**(12)** 3 3 3 3	Power point presentations / Lectures
Module II: Security, Compliance, Auditing, And Protection : • Steps to secure big data, Classifying Data • Protecting, Big Data Compliance Intellectual Property Challenge • Research Questions in Cloud Security, Open Problems	**(12)** **4** 4 4	Case Studies / Review of research articles
Module III: Hadoop Security Design: • Kerberos, Default Hadoop Model without security, • Hadoop Kerberos Security Implementation & Configuration.	**(12)** 6 6	Assignments
Module IV: Hadoop Ecosystem Security : • Configuring Kerberos for Hadoop ecosystem components – Pig, Hive, Oozie, Flume, HBase, Sqoop	**(12)** 12	Lectures
Module V: Data Security & Event Logging : • Integrating Hadoop with Enterprise Security Systems, • Securing Sensitive Data in Hadoop, • SIEM system, Setting up audit logging in Hadoop cluster.	**(6)** 4 4 4	Assignments / Lectures

Learning Resources	
1.	Mark Van Rijmenam, “Think Bigger: Developing a Successful Big Data Strategy for Your Business”, Amazon, 1 edition, 2014
2.	Frank Ohlhorst John Wiley & Sons, “Big Data Analytics: Turning Big Data into Big Money”, John Wiley & Sons, 2013.

3.	Sherif Sakr, “Large Scale and Big Data: Processing and Management”, CRC Press, 2014
4.	Ben Spivey, Joey Echeverria, “Hadoop Security Protecting Your Big Data Problem”, O’Reilly Media, 2015

REINFORCEMENT LEARNING

Credits: 4 **Semester: III**
Course Code: MDS23304B **No of Lecture Hours: 60**
Prerequisites: Machine Learning

Course Objective:
- Knowledge on fundamentals of reinforcement learning and the methods used to create agents that can solve a variety of complex tasks.

Course Outcome:
CO1: Understand basics of Monte-Carlo Method.
CO2: Understand RL model free methods.
CO3: Understand mathematical tricks that improve the performance of Temporal Difference
CO4: Acquire knowledge in finding the value of a state or an action when similar circumstances occur.
CO5: Analyze meaningful information from digital images, videos and other visual inputs.

Course Content	Hours Allotted	Pedagogy
Module I: Monte-Carlo methods: • Monte-Carlo methods: policy evaluation, • Rollouts, on policy and off-policy learning, • Importance sampling	**(12)** 4 4 4	Power point presentations / Lectures
Module II: Temporal Difference Learning : • Temporal Difference Learning: TD prediction, • Optimality of TD (0), SARSA, • Q-learning, Games and after states, • Maximization Bias and Double Learning.	**(12)** 4 4 4 4	Case Studies / Review of research articles
Module III: Eligibility Traces: • Eligibility traces: n-step TD prediction, • TD (lambda), forward and backward views, • Q(lambda), SARSA (lambda), Replacing traces and accumulating traces.	**(12)** 4 4 4	Assignments
Module IV: Function Approximation: • Function Approximation: Value prediction, • gradient descent methods, linear function approximation, Control algorithms, Fitted Iterative Methods, Deep Q-learning.	**(12)** 4 8	Lectures
Module V: Computer Vision: • Policy Gradient methods: non-associative learning - REINFORCE algorithm, • Exact gradient methods, estimating gradients, approximate policy gradient algorithms, actor-critic methods, Asynchronous Advantage Actor-Critic. • Hierarchical RL: MAXQ framework, Options framework, HAM framework, Option discovery algorithms. • Case studies: Elevator dispatching, Samuel's checker player, TD- gammon, Acrobot, Helicopter piloting, Alpha Go	**(12)** 4 4 4	Assignments / Lectures

Learning Resources	
1.	R. S. Sutton and A. G. Barto. Reinforcement Learning - An Introduction. MIT Press. 2nd Edition. 2018
2.	Video Lectures by Prof. David Silver
3.	Video Lectures by Prof. B.Ravindran

SCALABLE ARCHITECTURES FOR ML APPLICATIONS

Credits: 4 **Semester: III**
Course Code: MDS23304C **No of Lecture Hours: 60**

Course Objective:
- Students will be able to learn application and build Scalable Machine Learning
- Understand the students with Hadoop, SMACK Stack and Message Services.
- Be able to select the appropriate architecture for enterprise architecture based on the size, scale and applications used in the enterprise.

Course Outcome:
CO1: Understand the basic concepts of Scalable Machine Learning.
CO2: To become a data scientist work in some development environment tailored for statistics and Machine Learning.
CO3: Obtain expertise to turn actionable insights and Fast Data Applications into innovative methods to solve real-world problems.
CO4: Demonstrate the graph algorithms and live streaming data in Spark.
CO5: To impart knowledge on Kubernetes and batch processing.

Course Content	Hours Allotted	Pedagogy
Module I: • Introduction to Scalable Machine Learning, • Some Machine Learning Background Algorithms for Large scale Learning, • Overview of Hadoop and Current Big Data Systems	**(12)** 4 4 4	Power point presentations / Lectures
Module II: • How Programming for Data Flow Differs, • Basic Spark, Working with Vectors and Matrices in Spark, • Brief tour of Spark ML, • beyond parallelization, Practical Big Data	**(12)** 4 4 4 4	Case Studies / Review of research articles
Module III: • Anatomy of Fast Data Applications, • SMACK Stack – Functional Decomposition, • Message Backbone- Understanding messaging requirements, Data ingestion, Fast data& low latency, Message Delivery Semantics, Distributing Messages	**(12)** 3 3 6	Assignments
Module IV: • Computer Engines- Micro Batch Processing, • One-at-a time Processing, Choice of processing engine, • Storage as the Fast Data Borders, • The Message backbone as Transition Point	**(12)** 3 3 3 3	Lectures
Module V: • Sharing stateful streaming state, Data Driven Micro-services, • State and Micro-services. Deployment environments for Fast Data Applications, Application containerization,	**(12)** 4 4	Assignments / Lectures

• Resource scheduling, Apache Mesos, Kubernetes, Cloud Deployments.	4	

Learning Resources	
1.	Designing Fast Data Application Architectures by Gerard Maas, Stavros Kontopoulos, Sean Glover , Publisher: O'Reilly Media, Inc., June 2018
2.	Spark- The definitive Guide by Bill Chambers & Matei Zaharia, O'Reilly Media, Inc., June 2019

INTERNET OF THINGS

Credits: 4 **Semester: III**
Course Code: MDS25305 **No of Lecture Hours: 60**

Course Objective:
- To introduce the terminology, technology, and its applications
- To introduce the concept of M2M (machine to machine) with necessary protocols
- To introduce the deployment of IOT design methodology
- To introduce the Raspberry PI platform, that is widely used in IoT applications.
- To introduce the implementation of web-based services on IoT devices

Course Outcome:
CO1: Identify the importance of IOT and its applications.
CO2: Differentiate between IOT and M2M, SDN and NFV
CO3: Apply IOT design methodology.
CO4: Understand building of IOT devices and Raspberry PI.
CO5: Explain working of application of IOT.

Course Content	Hours Allotted	Pedagogy
Module I: Introductiona nd concepts: • Introduction to Internet of Things –Definition and Characteristics of IoT • Physical Design of IoT, Logical Design of IoT • IoT Enabling Technologies • IoT Levels and Deployment Templates, Domain Specific IoTs – Home Automation Cities, Environment, Agriculture, Industry, health, and Lifestyle	**(12)** 3 3 3 3	Power point presentations / Lectures
Module II: IoT and M2M: • IoT and M2M- Introduction to M2M, Difference between IoTandM2M • SDN and NFV for IoT **IoT System Management with Netconf-Yang** • Need for IoT Systems Management, SNMP • Network Operator requirements, NETCONF, YANG • IoT Systems Management with NETCONF-YANG	**(12)** 3 3 2 2 2	Case Studies / Review of research articles
Module III: IoT Design Methodology: • Purpose & Requirements Specification • Process Specification • Domain Model Specification • Information Model Specification • Service Specifications • IOT Level Specification • Functional View Specification • Operational View Specification • Device & Component Integration	**(12)** 1 1 1 1 1 1 1 1 1	Assignments

• Application Development • Weather monitoring system case study	1 2	
Module IV: Exploring IoT Data and Cloud Storage: • Cloud security and analytics • Applying big data technology to storage • Exploring and visualizing data • Techniques to understand data quality • Basic time series analysis • Statistical analysis	**(12)** 2 2 2 2 2 2	Lectures
Module V: Data Science for IoT Analytics: • Introduction to analytics and IoT analytics challenges • Introduction to Machine Learning • Machine Learning Methods • Feature engineering with IoT data • Validation methods • Understanding the bias–variance tradeoff	**(12)** 2 2 2 2 2 2	Assignments / Lectures

Learning Resources	
1.	Bahga, Arshdeep and Madisetti, Vijay. 2015. Internet of Things - A Hands-on Approach. Universities Press ISBN:9788173719547
2.	Richardson, Matt and Wallace, Shawn. 2014. Getting Started with RaspberryPi. O'Reilly (SPD). ISBN: 9789350239759
3.	Internet of Things Projects with Esp32: Build exciting and powerful IoT projects, 1st Edition, by AgusKurniawan

DEEP LEARNING

Credits: 4 **Semester: III**
Course Code: MDS25306 **No of Lecture Hours: 60**
Prerequisites: Machine Learning

Course Objective:
- To build the foundation of neural networks.
- Explore advanced topics of neural networks.
- To design and analyze various algorithms and techniques with a modern outlook.
- To enable the students, develop successful deep learning projects.

Course Outcome:
CO1: Learn the fundamental principles of deep learning.
CO2: Identify the deep learning algorithms for various types of learning tasks in various domains.
CO3: To explore Deep learning techniques and various feature extraction strategies.
CO4: To mathematically understand the deep learning approaches and paradigms.
CO5: Implement deep learning algorithms and solve real-world problems.

Course Content	Hours Allotted	Pedagogy
Module I: Introduction: • Deep Learning: Introduction, Historical trends and Challenges motivating Deep Learning • Supervised Learning Algorithms, Unsupervised Learning Algorithms • Linear Algebra: Scalars, Vectors, Matrices and Tensors • Gradient Based Optimization, Optimizing Deep Learning Models: Batch and Mini-batch Algorithms • Challenges in Neural Network Optimization: local minima, Plateaus and Saddle Points • Stochastic gradient descent, Momentum, and Nesterov Momentum method	**(12)** 2 2 2 2 2 2	Power point presentations / Lectures
Module II: Neural Nerworks: • Feedforward neural networks • Deep Networks • Architecture Design • Regularizing a Deep network • Model exploration and hyperparameter tuning.	**(12)** 2 2 2 2 4	Case Studies / Review of research articles
Module III: Convolution Neural Networks:: • Introduction to convolution neural networks • Convolution Operation • Motivation, Pooling • Convolution and Pooling as an Infinitely strong prior • Variants of the basic Convolution function • Structured outputs, Efficient Convolution Algorithms	**(12)** 2 2 2 2 2 2	Assignments
Module IV:Sequence Modeling : Recurrent Nets and	**(12)**	Lectures

Recursive Nets: • Unfolding computational graphs, recurrent neural networks (RNNs) • Bidirectional RNNs • Encoder-Decoder sequence to sequence architectures • Deep Recurrent networks • Recursive Networks • LSTM- The Long Short-Term Memory	2 2 2 2 2 2	
Module V: Auto encoders: • Under complete auto encoders • Regularized auto encoders, sparse autoencoders, denoising auto encoders • Representational power, layer, size, and depth of auto encoders • Stochastic encoders and decoders • Application of auto encoders	**(12)** 2 2 2 2 4	Assignments / Lectures

Learning Resources	
1.	Deep Learning by Ian Goodfellow, YoshuaBengio. Aaron Courville. The MIT Press, 2016
2.	Deep Learning, A Practical Approach by Rajiv Chopra, Khanna Book Publishing, 2018
3.	Jeff Heaton, Deep Learning and Neural Networks, Heaton Research Inc, 2015

NATURAL LANGUAGE PROCESSING

Credits: 4 **Semester: III**
Course Code: MDS25307 **No of Lecture Hours: 60**
Prerequisites: Python, Machine Learning

Course Objective:

- The goal is to become familiar with the concepts of the study of human language from a computational perspective.
- It covers syntactic, semantic and discourse processing models, emphasizing machine learning concepts.

Course Outcome:

CO1: Understanding Linguistic essentials and Applications of NLP.
CO2: Understand various approaches on Syntax Analysis in NLP.
CO3: Apply various approaches on Semantic Analysis & Language Model in NLP.
CO4: Analyze various methodologies used in Machine Translation, machine learning techniques used in NLP.
CO5: Analyze Information Retrieval and Lexical Resources

Course Content	Hours Allotted	Pedagogy
Module I: Introduction to NLP: • Background, overview • NLP Applications, NLP hard, Algorithms and Models, • Knowledge Bottlenecks in NLP, Introduction to NLTK • Linguistic Essentials, Chatbot Case Study	**(12)** 3 3 3 3	Power point presentations / Lectures
Module II: Word Level Analysis: • Regular Expressions, Text Normalization, Edit Distance • Syntax Analysis: Parsing, Spelling Error, Detection and Correction • Words and Word classes: Parts of Speech Tagging, Named Entity Recognition • Naive Bayes and Sentiment Classification: Case Study	**(12)** 4 3 2 3	Case Studies / Review of research articles
Module III: **Smoothed Estimation and Language Modelling:** • N-gram Language Models • Evaluating Language Models-The Language Modelling Problem • Smoothing: Add-1, Interpolation **Semantic Analysis and Discourse Processing** • Semantic Analysis: Elements, building blocks, Meaning Representation • Word Sense Disambiguation: Evaluation, Approaches, Applications, Drawbacks	**(12)** 3 2 3 2 2	Assignments
Module IV: Natural language Generation and Machine Translation: • Natural Language Generation: Architecture of NLG Systems,	**(12)** 3	Lectures

Applications • Machine Translation: Problems in Machine Translation, Machine Translation Approaches, Evaluation of Machine Translation systems. • Case study: Characteristics of Indian Languages	 4 5	
Module V: Information Retrieval and Lexical Resources • Information Retrieval: Design features of Information Retrieval Systems • Classical, Non- classical, Alternative Models of Information Retrieval, • Evaluation Lexical Resources: Word Embeddings, Word2vec, Glove. • Recommended Systems, Long short term memory (LSTM) • Linear Discriminant Analysis (LDA)	**(12)** 2 2 3 3 2	Assignments / Lectures

Learning Resources	
1.	Speech and Language Processing, DanielJurafskyandJamesH.,2nd Edition, Martin PrenticeHall,2013(UNIT-I, II, III, IV, V)
2.	Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press, 1999. (UNIT-I)
3.	Foundations of Computational Linguistics: Human-computer Communication in Natural Language, Roland R. Hausser, Springer, 2014.
4.	Other Electronic resources: https://www.intechopen.com/online-first/research-design-and-methodology
5.	Steven Bird, Ewan Klein and Edward Loper Natural Language Processing with Python, O'Reilly Media; 1 edition,2009.
6	https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf https://nptel.ac.in/courses/106101007/ NLTK – Natural Language Tool Kit http://www.nltk.org

INTERNET OF THINGS LAB (LAB VII)

Credits: 2 **Semester: III**
Course Code: MDS21308 **No of Lecture Hours: 60**

Course Objective:
- To Acquire the skill of IoT practical concepts.

Course Outcome:
- To Implement IoT based Practicals.

Lab Exercises

1. Setting up the Hardware and Software environment and library installation
2. Interfacing output devices
    a. LED
    b. Relay
    c. DCmotor
3. Interfacing input devices
    a. Switch
    b. IRsensor
4. Interfacing 16x2LCD
    a. Displaying Message
    b. Scrolling messages on LCD
    c. LM 35
5. Working with analog read and analog write functions.
    a. Reading data from Potentiometer and displaying it on LCD
    b. Interfacing Servo Motor and controlling using Potentiometer.
6. Interfacing Ultrasonic sensor for safety parking system and creating an excel dataset.
7. Interfacing DHT sensor and recording the values and creating an excel dataset.
8. Working on different API for communication – ThingSpeak or Ubidots for Desktop platform and BLYNK for mobile platform
9. Recording data from DHT, IR and Ultrasonic sensor onto ThinkSpeak or Ubidots and BLYNK
10. Controlling output devices from cloud platforms.

DEEP LEARNING LAB (LAB VIII)

Credits: 2 **Semester: III**
Course Code: MDS23309 **No of Lecture Hours: 60**

Course Objective:

- To Build The Foundation Of Deep Learning.
- To Understand How To Build The Neural Network.
- To enable students to develop successful machine learning concepts.

Course Outcome:

- Implement deep neural networks to solve real world problems.
- Choose appropriate pre-trained model to solve real time problem.
- Interpret the results of two different deep learning models.

Lab Exercises

1. Implement multilayer perceptron algorithm for MNIST Hand written Digit Classification.
2. Design a neural network for classifying movie reviews (Binary Classification) using IMDB dataset.
3. Design a neural Network for classifying news wires (Multi class classification) using Reuters dataset.
4. Design a neural network for predicting house prices using Boston Housing Price dataset.
5. Build a Convolution Neural Network for MNIST Hand written Digit Classification.
6. Build a Convolution Neural Network for simple image (dogs and Cats) Classification
7. Use a pre-trained convolution neural network (VGG16) for image classification.
8. Implement one hot encoding of words or characters.
9. Implement word embedding for IMDB dataset.
10. Implement a Recurrent Neural Network for IMDB movie review classification problem.

TEXT BOOKS:

Reza Zadeh and BharathRamsundar, “Tensorflow for Deep Learning”, O‟Reilly publishers, 2018

REFERENCES:

https://github.com/fchollet/deep-learning-with-python-notebooks

NATURAL LANGUAGE PROCESSING LAB (LAB –IX)

Credits: 2 **Semester: III**

Course Code: MDS25310 **No of Lecture Hours: 60**

Course Objective:
To acquire the skill of methodologies used in Machine Translation, Machine learning techniques used in NLP including unsupervised models and to analyze real time applications.

Course Outcome:
Implement the methodologies used in Machine Translation, machine learning techniques used in NLP including unsupervised models and to analyze real time applications.

Lab Exercises:

1. Import and display words from brown, book, wordnet corpus
2. Perform 'searching text' from corpus and display results
3. Count and display how often a word occurs in a text and plot same
4. Exercises on regular expressions
5. Implement and demonstrate the tokenize text.
6. Implement and demonstrate the count word frequency and to remove stop words.
7. Implement and demonstrate the tokenize Non-English Languages
8. Implement and demonstrate the get synonyms from WordNet.
9. Implement and demonstrate to get Antonyms from WordNet.
10. Implement and demonstrate the non-English Words
11. Implement and demonstrate the lemmatizing words Using WordNet.
12. Implement and demonstrate the different stemming and lemmatization words.
13. Implement and demonstrate the POS Tagging or Word Embeddings.
14. Case study-based program on Sentiment Analysis

}

FOURTH SEMESTER SCHEME

YEAR-WISE AND SEMESTER-WISE DISTRIBUTION OF SUBJECTS

M.SC. DATA SCIENCE FOURTH SEMESTER

ACADEMIC YEAR 2025-26 OF 2024-2026 BATCH (CBCS)

Sl. No.	Subject Code	TitleoftheSubject	Hours /week	Marks			Credits
				Internal	External	Total	
1	MDS21401	**MajorProject**	30	50	150	200	12
Total			**30**	**50**	**150**	**200**	**12**

PROJECT

Credits: 12 **Semester: IV**
Course Code: MDS21401 **No of Lecture Hours: 90**

Course Objective:

- The primary emphasis of the project work is to understand and gain knowledge of the principles of data science and analytics practices.

Course Outcome:

- The student will be able to describe Data Collection, Data Visualization, Labeling, Data Selection, Data Pre-processing, Data transformation, Model Training, Model evaluation, Model testing, Accuracy, and deployment.

Course Content	Hours Allotted	Pedagogy
Module I: Title: • Fourth Semester of the M.Sc. Data Science course is exclusively meant for project work. Project must be carried out by each student individually in a period of 15 weeks. • Students should submit a synopsis at the end of 3rd Semester in consultation with the Project Guide. • The synopsis should consist of Data Collection, DataVisualization, Labeling, Data Selection, DataPre-processing, Data transformation, Model Training, Model evaluation, Model testing, Accuracy and deployment are required to show in the project work. • A committee consisting of the department faculty members of the respective college alongwith a guide will evaluate the project and award internal marks. • At the end of the semester the students are required to present themselves for aViva-voce examination in which each student will be awarded an external mark.	90	**Case Studies/Presentations**