Syllabus

MSc (Data Science)

CHRIST (Deemed to be University), Bangalore.

Karnataka, India

AY 2024-25

Index

Department Overview:

The Department of Statistics and Data Science, established in the year 2022, strives to provide a dynamic research environment and effective education, including excellent training in scientific data collection, data management, methods and procedures of data analysis. Our, curriculum adheres to worldwide standards to provide the best possible research and educational/Industry opportunities.

It offers a perfect blend of statistical knowledge with tools and data science techniques required to explode, analyze and interpret the complex data of the modern world. The curriculum and teaching pedagogy foster higher-order thinking and research skills, which equip students for the dynamic and ever-evolving data industry. Well-designed co-curricular activities organized by the department are aimed at the holistic development of students. The skills imparted through various programs offered by the department help in interdisciplinary research for the benefit of the society.

Vision and Mission:

Vision:

Excellence and Service

Mission:

To develop statistics and data science professionals capable of enriching sustainable and progressive society for achieving common national goals.

Programme Description:

Data Science is popular in all academia, business sectors, and research and development to make effective decision in day to day activities. MSc in Data Science is a two year programme with VI Trimester. This programme aims to provide opportunity to all candidates to master the skill sets specific to data science with research bent. The curriculum supports the students to obtain adequate knowledge in theory of data science with hands on experience in relevant domains and tools. Candidate gains exposure to research models and industry standard applications in data science through guest lectures, seminars, projects, internships, etc.

Programme Outcomes

PO1: Problem Analysis and Design: Ability to identify analyze and design solutions for data science problems using fundamental principles of mathematics, Statistics, computing sciences, and relevant domain disciplines.

PO2: Enhance disciplinary competency and employability: Acquire the skills in handling data science programming tools towards problem solving and solution analysis for domain specific problems.

PO3: Societal and Environmental Concern: Utilize the data science theories for societal and environmental concerns

PO4: Professional Ethics: Understand and commit to professional ethics and professional computing practices to enhance research culture and uphold the scientific integrity and objectivity

PO5: Individual and Team work: Function effectively as an individual and as a member or leader in diverse teams and in multidisciplinary environments.

PO6: Engage in continuous reflective learning in the context of technology advancement: Understand the evolving data and analysis paradigms and apply the same to solve the real life problems in the fields of data science.

Programme Eligibility:

A candidate who has passed an Undergraduate degree with 50 % aggregate marks from any University in India or abroad that is recognized by UGC / AIU. Students must fulfill either criteria A or B described below in order to be eligible for the programme:

A. Bachelor of Computer Applications (BCA) / BSc Computer Science/ BSc Data Science / BE Computer Science
OR
B. BE/B Tech/Under Graduate degree in Science with any two of the following subjects as major or minor (minimum of One years of learning)
    1. Computer Science
    2. Mathematics
    3. Statistics

PROGRAMME STRUCTURE:

Trimester	No. of Course	Hours	Credits	Marks
Trimester 1	7	28	19	550
Trimester 2	5	29	18	500
Trimester 3	7	29	19	600
Trimester 4	6	29	16	550
Trimester 5	5	29	16	500
Trimester 6	2	6	12	350
Total			**100**	**3050**

TRIMESTER-I

TYPE	Course Code	Course Title	Course hrs	Hours Per Week	Credits	Marks
ESE	MDS131	Research methods in Data Science	60	5	4	100
ESE	MDS132	Probability and Distribution Theory	60	5	4	100
ESE	MDS133	Mathematical Foundations for Data Science-I	45	4	3	100
	Choose Any One (Foundational Elective)					
FULL CIA	MDS161A	Found tion Elective-I (Principles of Programming)	30	3	2	50
	MDS161B	Foundation Elective-II (Introduction to Probability and Statistics)				
	MDS161C	Foundation Elective-III (Linux Essentials)				
FULL CIA	MDS171	Programming using Python	75	7 (4+3)	4	100
FULL CIA	MDS151	Applied Excel	30	3	1	50
	HED	HOLISTIC EDUCATION		1	1	50
	Total	-		**28**	**19**	**550**

TRIMESTER-II

TYPE	Course Code	Course Title	Course hrs	Hours Per Week	Credits	Marks
ESE	MDS231	Design and Analysis of Algorithms	45	4	3	100
ESE	MDS232	Mathematical Foundations for Data Science-II	45	4	3	100
FULL CIA	MDS271	Database Technologies	75	7 (4+3)	4	100
FULL CIA	MDS272	Inferential Statistics using R	75	7 (4+3)	4	100
FULL CIA	MDS273	Full Stack Web Development	75	7 (4+3)	4	100
		Total		**29**	**18**	**500**

TRIMESTER-III

TYPE	**Course Code**	**Course Title**	**Course hours**	**Hours Per Week**	**Credits**	**Marks**
ESE	MDS331	Regression Modeling	45	4	3	100
FULL CIA	MDS371	Java Programming	75	7 (4+3)	4	100
FULL CIA	MDS372	Machine Learning	75	7(4+3)	4	100
		ELECTIVE (Statistics - Concepts Based)				
ESE	MDS341A	Categorical Data Analysis	45	4	3	100
	MDS341B	Multivariate Analysis				
	MDS341C	Stochastic Processes				
FULL CIA	MDS381	SEMINAR	30	3	2	50
FULL CIA	MDS311	Cloud Services	30	3	2	100
	HED	HOLISTIC EDUCATION		1	1	50
		Total		**29**	**19**	**600**

TRIMESTER-IV

Type	Course Code	Course Title	Course hrs	Hours Per Week	Credits	Marks
ESE	MDS431	Data driven Modelling and Visualization	30	3	2	100
ESE	MDS432	Time Series and Forecasting Techniques	60	5	4	100
FULL CIA	MDS471	Neural Networks and Deep Learning	75	7 (4+3)	4	100
		ELECTIVES (Data Science)				
FULL CIA	MDS472A	Web Analytics	60	6 (3+3)	3	100
FULL CIA	MDS472B	IoT Analytics				
FULL CIA	MDS472C	Natural Language Processing				
FULL CIA	MDS472D	Graph Analytics				
	MDS481	PROJECT-I (Web project with Data Science concepts)	60	5	2	100
	MDS482	RESEARCH PROBLEM identification	30	3	1	50
		Total		**29**	**16**	**550**

TRIMESTER-V

TYPE	Course Code	Course Title	Course hrs	Hours Per Week	Credits	Marks
FULL CIA	MDS571	Big Data Analytics	75	7 (4+3)	4	100
		ELECTIVE - 1 (Applied Statistics)				
ESE	MDS531A	Econometrics	60	5	4	100
ESE	MDS531B	Bayesian Inference				
ESE	MDS531C	Bio- statistics				
		ELECTIVE-2 (Emerging analysis paradigms)				
FULL CIA	MDS572A	Evolutionary Algorithms	60	6 (3+3)	3	100
FULL CIA	MDS572B	Quantum Machine Learning				
FULL CIA	MDS572C	Reinforcement Learning				
FULL CIA	MDS573A	Geospatial Data Analytics	60	6 (3+3)	3	100
FULL CIA	MDS573B	Bio-Informatics				
FULL CIA	MDS573C	Image and Video Analytics				
	MDS581	Project - II (Research Project/ Data Science Capstone Project)	60	5	2	100
		Total		**29**	**16**	**500**

TRIMESTER-VI

Type	Course Code	Course Title	Course hrs	Hour s Per Week	Credits	Marks
	MDS681	Industry Project	30	3	10	300
	MDS682	RESEARCH PUBLICATION	30	3	2	50
		Total		**6**	**12**	**350**

MDS 131: RESEARCH METHODS IN DATA SCIENCE

Total Teaching Hours for Trimester: 60
No of hours per week: 5
Max Marks: 100 **Credits: 4**
Course Type: Major Course

Description
To assist students in planning and carrying out research work in the field of data science. The students are exposed to the basic principles, procedures and techniques of implementing a research project. The course provides a strong foundation for data science and the application area related to it. Students are trained to understand the underlying core concepts and the importance of ethics while handling data and problems in data science.

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

No.	Course Outcomes
CO1	Understand the essence of research and the importance of research methods and methodology
CO2	Explore the fundamental concepts of data science
CO3	Understand various machine learning algorithms used in data science process
CO4	Learn to think through the ethics surrounding privacy, data sharing and algorithmic decision- making
CO5	Create scientific reports according to specified standards

Unit 1 **Teaching Hours: 12**
Research Methodology
Introduction: Objectives of Research, Types of Research, Research Approaches, Significance of Research, Research Methods versus Methodology. Defining research problem: Selecting the problem, Necessity of defining the problem, Techniques involved in defining a problem, Research Design: Different Research Designs, Basic Principles of Experimental Designs, Developing a Research Plan.

Unit 2 **Teaching Hours: 12**
Introduction to Data Science

Definition – Big Data and Data Science Hype – Why data science – Getting Past the Hype – The Current Landscape – Who is a Data Scientist? - Data Science Process Overview – Defining goals – Retrieving data– Data preparation – Data exploration – Data modeling– Presentation.

Sampling, Measurement and Scaling Techniques Sampling: Steps in Sampling Design, Different Types of Sample Designs, Measurement and Scaling: Measurement in Research, Measurement Scales, Technique of Developing Measurement Tools, Scaling, Important Scaling Techniques.

Unit 3 **Teaching Hours: 12**

Machine Learning

Machine learning – Modeling Process – Training model – Validating model – Predicting new observations – Supervised learning algorithms – Unsupervised learning algorithms.

Unit 4 **Teaching Hours: 12**

Report Writing

Working with Literature: Importance, finding literature, Using the resources, Managing the literature, Keep track of references, Literature review. Scientific Writing and Report Writing: Significance, Steps,Layout, Types, Mechanics and Precautions**,**
Latex: Introduction, Text, Tables, Figures, Equations, Citations, Referencing, and Templates (IEEE style), Paper writing for international journals, Writing scientific report.

Unit 5 **Teaching Hours: 12**

Ethics in Research and Data Science

Research ethics, Data Science ethics – Doing good data science – Owners of the data - Valuing different aspects of privacy - Getting informed consent - The Five Cs –Diversity – Inclusion.

Essential Reading

1. Davy Cielen and Arno Meysman, Introducing Data Science. Simon and Schuster, 2016.
2. M. Loukides, H. Mason, and D. Patil, Ethics and Data Science. O’Reilly Media, 2018.
3. C. R. Kothari, Research Methodology Methods and Techniques. 3rd. ed. New Delhi: NewAge International Publishers, Reprint 2014.
4. Zina O’Leary, The Essential Guide of Doing Research. New Delhi: PHI, 2005

Recommended Reading

1. Data Science from Scratch: First Principles with Python, Joel Grus, O’Reilly, 1st edition,2015
2. Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt, O’Reilly, 1st edition, 2013
3. Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Cambridge University Press, 2nd edition, 2014
4. 4.Sinan Ozdemir, Principles of Data Science learn the techniques and math you need tostart making sense of your data. Birmingham Packt December, 2016.
5. J. W. Creswell, Research Design: Qualitative, Quantitative, and Mixed MethodsApproaches. 4thed. SAGE Publications, 2014.
6. Kumar, Research Methodology: A Step-by-Step Guide for Beginners. 3rd. ed. Indian: PE,2010.

MDS132: Probability and Distribution Theory

Total Teaching Hours for Semester: 60
No of hours per week: 5
Max Marks: 100 **Credits: 4**
Course Type: Major

Course Description

Probability and probability distributions play an essential role in modeling data from the real-world phenomenon. This course will equip students with thorough knowledge in probability and various probability distributions and model real-life data sets with an appropriate probability distribution.

Course Outcomes

Upon completion of the course, students will be able to

No.	Course Outcomes
CO1	able to understand the concept of the random variable and expectation for discrete and continuous data
CO2	evaluate condition probabilities and conditional expectations
CO3	gain the knowledge of applications of discrete distributions in Data Science
CO4	identify the applications of continuous distributions in Data Science
CO5	apply Chebychevs inequality to verify the convergence of sequence in probability

Unit 1 **Teaching Hours: 12**
Random Variables and Expectations

Random Variables: Definitions and Properties, Distribution Function and its properties, Discrete and Continuous Random Variables; Expectations: Expected value of a random variable, Properties of expectation, Moment Generating Function Use of moments for mean, variance and moments.

Unit 2 **Teaching Hours: 12**
Joint Distributed Random Variables

Joint probability mass function, joint probability density function, joint distribution functions, marginal functions, conditional distribution functions, conditional pdf, conditional pmf, joint moments, covariance, correlation, conditional expectation.

Unit 3 **Teaching Hours: 12**
Probability Distribution for Discrete Data

Bernoulli, Binomial, Poisson, Negative Binomial, Hypergeometric Distributions (mean and variance in terms of mgf), their applications in Data Science.

Unit 4 **Teaching Hours: 12**

Probability Distribution for Continuous Data
Uniform, Normal, Exponential, Gamma distribution, Weibull Distributions (mean and variance in terms of mgf), and their applications in Data Science.

Unit 5 **Teaching Hours: 12**
Limit Theorems
Chebychev's inequality - weak law of large numbers (iid): examples - strong law of large numbers (statement only) - central limit theorems (iid case): examples.

Essential Reading

1. Introduction to theory of Statistics. A.M. Mood, F.A. Graybill & D.C. Boes, Tata McGraw Hill, 3rd Edition, 2017.
2. Introduction to Probability Models. S.M. Ross, Academic Press, 12th Edition, 2019.
3. Fundamentals of Mathematical Statistic. S.C. Gupta, & V.K. Kapoor, Sultan Chand & Sons Publications, 12th Edition, 2022

Recommended Reading

1. A first course in Probability. S.M. Ross, Pearson, 10th Edition, 2019.
2. An Introduction to Probability and Statistics. V.K. Rohatgi & A.K.Md.E. Saleh, Wiley, 3rd Edition, 2015.

MDS161A: Principles of Programming

Total Teaching Hours for Semester: 30
No of hours per week: 03
Max Marks: 50 **Credits: 2**
Course Type: Foundational Elective

Course Objectives
The students shall be able to understand the main principles of programming. The objective also includes indoctrinating the activities of implementation of programming principles.

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

No.	Course Outcomes
CO1	Understand the fundamentals of programming languages.
CO2	Understand the design paradigms of programming languages.
CO3	To examine expressions, subprograms and their parameters.

Unit 1 **Teaching Hours: 10**
Introduction
Introduction to Syntax and Grammar : Introduction, Programming Languages, Syntax, Grammar, Ambiguity, Syntax and Semantics, Data Types (Primitive/Ordinal/Composite data types, Enumeration and sub-range types, Arrays and slices, Records, Unions, Pointers and pointer problems)

Unit 2 **Teaching Hours: 10**
Constructing Expressions
Expressions, Type conversion, Implicit/Explicit conversion, type systems, expression evaluation, Control Structures, Binding and Types of Binding,Lifetime,Referencing Environment (Visibility,Local/Nonlocal/Global variables), Scope (Scope rules, Referencing operations, Static/Dynamic scoping).

Unit 3 **Teaching Hours: 10**
Subprograms and Parameters
Subprograms, signature, Types of Parameters, Formal/Actual parameters, Subprogram overloading, Parameter Passing Mechanisms, Aliasing, Eager/Normal-order/Lazy evaluation) , Subprogram Implementation (Activation record, Static/Dynamic chain, Static chain method, Deep/Shallow access, Subprograms as parameters, Labels as parameters, Generic subprograms, Separate/Independent compilation).

Essential Reading

1. Allen B. Tucker, Robert Noonan, Programming Languages: Principles and Paradigms, Tata McGraw Hill Education, 2006.
2. Bruce J. MacLennan, “Principles of Programming Languages: Design, Evaluation, and Implementation”, Third Edition, Oxford University Press (New York), 1999.

Recommended Reading

1. T. W. Pratt, M. V. Zelkowitz, Programming Languages, Design and Implementation, Prentice Hall, Fourth Edition, 2001
2. Robert Harper, Practical Foundations for Programming Languages, Second Edition, Cambridge University Press, 2016.

MDS161B: INTRODUCTION TO PROBABILITY AND STATISTICS

Total Teaching Hours forSemester:30
No of hours per week: 3
Max Marks: 50 **Credits: 2**
Course Type: Foundation Elective Course

Description
This course is designed to introduce the historical development of statistics, presentation of data, descriptive measures and cultivate statistical thinking among students. This course also introduces the concept of probability.

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

No.	Course Outcomes
CO1	Demonstrate, present and visualize data in various forms, statistically.
CO2	Understand and apply descriptive statistics.
CO3	Evaluation of probabilities for various kinds of random events.

Unit 1 **Teaching Hours: 8**
ORGANIZATION AND PRESENTATION OF DATA
Origin and development of Statistics - Scope - limitation and misuse of statistics - types of data: primary, secondary, quantitative and qualitative data - Types of Measurements: nominal, ordinal, ratio and scale - discrete and continuous data - Presentation of data by tables - graphical representation of a frequency distribution by histogram and frequency polygon - cumulative frequency distributions (inclusive and exclusive methods).

Unit 2 **Teaching Hours: 6**
DESCRIPTIVE STATISTICS I
Measures of location or central tendency: Arithmetic mean - Median - Mode - Geometric mean - Harmonic mean.

Unit 3 **Teaching Hours: 6**
DESCRIPTIVE STATISTICS II
Partition values: Quartiles - Deciles and Percentiles - Measures of dispersion: Mean deviation - Quartile deviation - Standard deviation - Coefficient of variation- Moments: measures of skewness - kurtosis.

Unit 4 **Teaching Hours: 10**
BASICS OF PROBABILITY
Random experiment - sample point and sample space– event - algebra of events - Definition of Probability: classical - empirical and axiomatic approaches to probability - properties of

probability - Theorems on probability - conditional probability and independent events - Laws of total probability - Baye's theorem and its applications.

Essential Reading

1. David C. Lay, Steven R. Lay, Judi J. McDonald (2016) Linear algebra and its applications. Pearson. S. Axler, Linear algebra done right, Springer, 2017
2. Strang, G. (2006) Linear Algebra and its Applications: Thomson Brooks. Cole, Belmont, CA, USA

Recommended Reading

1. E. Davis, Linear algebra and probability for computer science applications, CRC Press, 2012
2. J. V. Kepner and J. R. Gilbert, Graph algorithms in the language of linear algebra, Society for Industrial and Applied Mathematics, 2011.
3. D. A. Simovici, Linear algebra tools for data mining, World Scientific Publishing, 2012.
4. P. N. Klein, Coding the matrix: linear algebra through applications to computer science, Newtonian Press, 2015.

MDS161C: LINUX ADMINISTRATION

Total Teaching Hours forSemester:30
No of hours per week: 3
Max Marks: 50 **Credits: 2**
Course Type: Foundational Elective

Course Description
This course is designed to introduce the Linux working environment to students. This course will enable students to understand the Linux system architecture, File and directory commands and foundations of shell scripting.

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

No.	Course Outcomes
CO1	Demonstrate the Basic file, directory commands
CO2	Understand the Unix system environment
CO3	Apply shell programming concepts to solve given problem

Unit 1 **Teaching Hours: 10**
Introduction
Introduction, Salient features, Unix system architecture,Unix Commands, Directory Related Commands, File Related Commands,Disk related Commands,General utilities,Unix File System,Boot inode, super and data block ,in core structure,Directories, conversion of path name to inode, inode to new file,Disk block allocation.

Unit 2 **Teaching Hours: 10**
Process Management
Process Management Process state and data structures of a Process,Context of a Process, background processes,User versus Kernel node,Process scheduling commands,. Process scheduling commands,Process terminating and examining commands,Secondary Storage Management: Formatting, making file system, checking disk space, mountable file system, disk partitioning.

Unit 3 **Teaching Hours: 10**
Shell Programming
Shell Programming, Vi Editor,.Shell types, Shell command line processin Shell script & its features, system and user defined variables, Executing sexpr command Shell Screen Interface, read and echo statement,Shell ScriConditional Control Structures – if statement,Case statement,Looping C while,for,Jumping Control Structures – break, continue, exit.

Essential Reading:

1. Linux: The Complete Reference, sixth edition, Richard Petersen, 2017

Recommended Reading:

1. Linux Pocket Guide, Daniel J. Barrett,3rd edition, O'Reilly

MDS171 - PROGRAMMING USING PYTHON

Total Teaching Hours for Trimester: 75
No of hours per week: 7(4+3)
Max Marks: 150 **Credits: 4**
Course Type: Major

Course Description
The objective of this course is to provide comprehensive knowledge of Python programming paradigms required for Data Science.

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

No.	Course Outcomes
CO1	Demonstrate the use of built-in objects of Python
CO2	Demonstrate significant experience with Python program development environment
CO3	Implement numerical programming, data handling and visualization through NumPy, Pandas and MatplotLib modules.

UNIT 1 **Teaching Hours: 15**
Introduction to Python
Python and Computer Programming - Using Python as a calculator - Python memory management - Structure of Python Program - Branching and Looping - Problem-Solving Using Branches and Loops - Lists and Mutability - Functions - Problem-Solving Using Lists and Functions.
Lab Exercise
1. Variables, constants and inbuilt functions.
2. Demonstrate usage of branching and looping statements
3. Demonstrate Recursive functions
4. Demonstrate Lists

UNIT 2 **Teaching Hours: 15**
Sequence Datatypes and Object-Oriented Programming
Sequences, Mapping and Sets - Dictionaries - Classes: Classes and Instances -Inheritance - Exceptional Handling - Module: Built-in modules & user-defined module - Introduction to Regular Expressions using “re” module
Lab Exercises
1. Demonstrate Tuples, Sets, Frozen sets and Dictionaries
2. Demonstrate inheritance and exception handling
3. Demonstrate the use of “re”

UNIT 3 **Teaching Hours: 15**

NUMPY

Basics of NumPy - Computation on NumPy - Aggregations - Computation on Arrays- Comparisons, Masks and Boolean Arrays -Sorting Arrays - Structured Data: NumPy's Structured Array.

Lab Exercises

1. Demonstrate Aggregation
2. Demonstrate Indexing and Sorting
3. Demonstrate handling of missing data
4. Demonstrate hierarchical indexing

UNIT 4 **Teaching Hours: 15**

Data Manipulation with Pandas

Introduction to Pandas Objects - Data indexing and Selection - Operating on Data in Pandas - Handling Missing Data - Hierarchical Indexing - Aggregation and Grouping - Pivot Tables - Vectorized String Operations - High-Performance Pandas: eval() and query().

Lab Exercises

1. Demonstrate usage of Pivot Table
2. Demonstrate use of eval() and query()

UNIT 5 **Teaching Hours: 15**

Visualization with Matplotlib

Basics of matplotlib - Simple Line Plot and Scatter Plot - Density and Contour Plots - Histograms, Binnings and Density - Customizing Plot Legends - Multiple subplots.

Lab Exercises

1. Demonstrate Line plot, Bar and Pie Chart.
2. Demonstrate Scatter Plot, Histogram, KDE, Violin Plot

Essential Reading:

1. Jake VanderPlas, Python Data Science Handbook - Essential Tools for Working with Data, O'Reily Media, Inc, 2016
2. Zhang. Y, An Introduction to Python and Computer Programming, Springer Publications, 2016

Recommended Reading:

1. JoelGrus, Data Science from Scratch First Principles with Python, O'Reilly, Media,2016
2. T. R. Padmanabhan, Programming with Python, Springer Publications, 2016.M. Rajagopalan and P. Dhanavanthan- Statistical Inference-1st ed. - PHI Learning (P) Ltd.- New Delhi- 2012.
3. V. K. Rohatgi and E. Saleh- An Introduction to Probability and Statistics- 3rd ed.- John Wiley & Sons Inc- New Jersey- 2015.

MDS151: Applied Excel

Total Teaching Hours/Trimester: 30
No. of Lecture Hours/Week: 3P
Maximum Marks: 50 **Credits: 1**
Course Type: Major

Course description:
This course is designed to build logical thinking ability and to provide hands-on experience in solving statistical models using MS Excel with Problem based learning. To explore and visualize data using excel formulas and data analysis tools.

Course Objective:
The course enables the students to work with different kinds of data into excel. The students can analyze, infer and visualize data using excel formulas and methods.

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

No.	Course Outcomes
CO1	Demonstrate the data management using excel features.
CO2	Analyze the given problem and solve using Excel.
CO3	Infer the building blocks of excel, excel shortcuts, sample data creation

Unit 1 **Teaching Hours: 10**
Layout
Introduction: File types - Spreadsheet structure - Menu bar - Quick access toolbar - Mini toolbar - Excel options - Formatting: Format painter - Font - Alignment - Number - Styles - Cells, Clear - Page layout. Properties Symbols - Equation - Editing - Link - Filter - Charts - Formula Auditing
Overview of Excel tables and properties - Collecting sample data and arranging in definite format in Excel tables.
Lab Exercises:
1. Excel Formulas
2. Excel Tables and Properties
3. Data Collection
4. Excel Charts

Unit 2 **Teaching Hours: 10**
Files
Importing data from different sources - Exporting data in different formats
Database
Creating database with the imported data - Data tools: text to column - identifying and removing duplicates - using format cell options.
Lab Exercises:
1. Import data

2. Export data
3. Creating database
4. Data tools

Unit 3 **Teaching Hours: 10**

Functions

Application of functions - Concatenate - Upper - Lower - Trim - Repeat - Proper - Clean - Substitute - Convert - Left - Right - Mid - Len - Find - Exact - Replace - Text join - Value - Fixed etc.

Lab Exercises:

1. Excel functions.

Essential Reading:

1. Alexander R, Kuselika R and Walkenbach J, Microsoft Excel 2019 Bible, Wiley India Pvt Ltd, New Delhi, 2018.

Recommended Reading

1. Paul M, Microsoft Excel 2019 formulas and functions, Pearson Eduction, 2019.

MDS231: Design and Analysis of Algorithms

Total Teaching Hours for Trimester: 45
No of hours per week: 4
Max Marks: 100 **Credits: 3**
Course Type: Major

Course Description The course studies techniques for designing and analyzing algorithms and data structures. It concentrates on techniques for evaluating the performance of algorithms. The objective is to understand different designing approaches like greedy, divide and conquer, dynamic
programming etc. for solving different kinds of problems.

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

No.	Course Outcomes
CO1	Design new algorithms and analyze their asymptotic and absolute runtime and memory demands.
CO2	Apply classical sorting, searching, optimization and graph algorithms.
CO3	Understand basic techniques for designing algorithms, including the techniques of recursion, divide-and-conquer, greedy algorithm etc.
CO4	Understand the mathematical criterion for deciding whether an algorithm is efficient and know many practically important problems that do not admit any efficient algorithms.

UNIT 1 **Teaching Hours: 9**
Introduction
Algorithms, Analyzing algorithms, Complexity of algorithms, Growth of functions, Performance measurements, Sorting and order Statistics - Shell sort, Sorting in linear time, Linear Search

UNIT 2 **Teaching Hours: 9**
Advanced Data Structures
Red-Black trees, B – trees, Binomial Heaps, Fibonacci Heaps, Tries, skip list.

UNIT 3 **Teaching Hours: 9**
Divide and Conquer
Quick sort, Merge sort,Matrix Multiplication Binary Searching. **Greedy methods** with examples such as Optimal Reliability Allocation, Knapsack, Minimum Spanning trees – Prim's and Kruskal's algorithms, Single source shortest paths - Dijkstra's algorithms.Optimal merge patterns.

UNIT 4 **Teaching Hours: 9**
Dynamic Programming

Dynamic programming with examples such as Knapsack, All pair shortest paths – Warshal's andFloyd's algorithms,Backtracking, n-Queen Problem, Sum of subsets,Graph Coluring, Branch and Bound with examples such as Travelling Salesman Problem .

UNIT 5 **Teaching Hours: 9**
Selected Topics
Algebraic Computation, Fast FourierTransform, String Matching, Theory of NP-completeness, Approximation algorithms and Randomized algorithms.

Essential Reading

1. Coreman, Rivest, Lisserson, "An Introduction to Algorithm", PHI, 2001
2. Horowitz & SAHANI," Fundamental of computer Algoritm", Galgotia Publications, 2nd Edition.

Recommended Reading

1. Aho, Hopcraft, Ullman, "The Design and Analysis of Computer Algorithms" Pearson Ed9ucation, 2008.
2. Donald E. Knuth, *The Art of Computer Programming Volume 3, Sorting and Searching*, 2nd Edition, Pearson Education, Addison-Wesley, 1998.
3. GAV PAI, *Data structures and Algorithms,* Tata McGraw Hill, Jan 2008.

MDS271: DATABASE TECHNOLOGIES

Total Teaching Hours for Semester: 75
No of hours per week: 7(4+3)
Max Marks: 100 **Credits: 4**
Course Type: Major

Course Description
The main objective of this course is fundamental knowledge and practical experience with database concepts. It includes the concepts and terminologies which facilitate the construction of relational databases, writing effective queries, comprehending data warehouse and NoSQL databases and its types.
Course Outcomes:
Upon completion of the course students will be able to

No.	Course Outcomes	LRNG Needs
CO1	Demonstrate various databases and compose effective queries	Global
CO2	Understanding the process of OLAP system construction	Global
CO3	Develop applications using Relational and NoSQL databases.	Global

UNIT 1 INTRODUCTION **Teaching Hours: 15**
Concept & Overview of DBMS, Data Models, Database Languages, Database Administrator, Database Users, Three Schema architecture of DBMS. Basic concepts, Design Issues, Mapping Constraints, Keys, Entity-Relationship Diagram
Lab Exercises
1. SQL-DML commands

UNIT 2 RELATIONAL MODEL AND DATABASE DESIGN **Teaching Hours: 15**
SQL and Integrity Constraints, Concept of DDL, DML, DCL. Basic Structure, Set operations, Aggregate Functions, Null Values, Domain Constraints, Referential Integrity Constraints, Assertions, Views, Nested Subqueries, Functional Dependency, Different anomalies in designing a Database, Normalization: using functional dependencies, Normal Forms. 1NF, 2 NF, 3NF, BCNF
Lab Exercises
1. SQL DML Statements 1
2. SQL DML Statements 2

UNIT 3 DATA WAREHOUSE: THE BUILDING BLOCKS **Teaching Hours: 15**
Defining Features, Database and Data Warehouses, Architectural Types, Overview of the Components, Metadata in the Data warehouse, The Star Schema, Star Schema Keys, Advantages of the Star Schema, Star Schema: Examples, Snowflake Schema, Aggregate Fact Tables.
Lab Exercises
1. SQL-DML Statements(Set operations, Joins, Sub Queries)
2. Dimensional Modelling for Data warehousing

UNIT 4 INTRODUCTION TO NOSQL DATABASES **Teaching Hours: 15**

Overview, and History of NoSQL Databases. Attack of the Clusters, the Emergence of NoSQL, Key Points comparison of relational databases to new NoSQL stores, RDBMS approach, Challenges NoSQL approach, Key-Value Data Model, Document Data Models, Column Family Stores, Graph Databases

Lab Exercices
1. MongoDB DB and Collections
2. MongoDB Inserts & Updates

UNIT 5 DOCUMENT DATABASE-MONGODB Teaching Hours: 15

Distributed Databases- Sharding and Replication, Consistency, The CAP Theorem.Document Data Model: Documents and Collections. Embedded Collection. **CRUD** (Creating, Reading & Updating Data) -Mongo Shell, Query Operators, Projection Operation, InsertOne, InsertMany Update Operators, and a Few Commands

Lab Exercices
1. MongoDB: Data manipulation and Searches
2. MongoDB: Data import from external sources

Essential Reading
1. Henry F. Korth and Silberschatz Abraham, "Database System Concepts", Mc.Graw Hill.
2. Thomas Cannolly and Carolyn Begg, "Database Systems, A Practical Approach to Design, Implementation and Management", Third Edition, Pearson Education, 2007.
3. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd John Wiley & Sons, Inc. New York, USA, 2002
4. Sadalage, P. & Fowler, M. (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Pearson Education

Recommended Reading
1. LiorRokach and OdedMaimon, Data Mining and Knowledge Discovery Handbook, Springer, 2nd edition, 2010
2. Redmond, E. & Wilson, J. R. (2012). Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement (1st ed.) O′Reilly.

MDS272: INFERENTIAL STATISTICS

Total Teaching Hours for Semester: 75
No of hours per week: 7(4+3)
Max Marks: 100 **Credits: 4**
Course Type: Major

Course Description
Statistical inference plays an important role when analyzing data and making decisions based on real-world phenomena. This course aims to teach students to test hypotheses and estimate parameters for real life data sets.
Course Outcomes: Upon completion of the course students will be able to

No.	Course Outcomes
CO1	Demonstrate the concepts of population and samples
CO2	Apply the idea of sampling distribution of different statistics in testing of hypothesis
CO3	Estimate the unknown population parameters using the concepts of point and interval estimations using R.
CO4	Test the hypothesis using nonparametric tests for real world problems using R.

CO-UNIT MAPPING:
UNIT 1: ESTIMATION TECHNIQUES **Teaching Hours: 15**
Population and Sample, Parameter and Statistics, Characteristics of a good estimators, Suffiency -Factorisation Theorem, Unbiased Estimators- Consistency, Efficiency, Different methods of estimation-Moment estimation and MLE Estimation techniques.

Lab Exercises:
1. Introduction to R, usage of R as a basic calculator.
2. Creating a vector, Accessing elements of the vector, matrix operations in R.
3. Calculation of sampling error and standard error, power of the test.
4. Simulation of random variable and its estimation.

UNIT 2: TESTING OF HYPOTHESIS I **Teaching Hours: 15**
Concept of large and small samples, Single sample mean test, Independent sample mean test, Paired sample mean test, Test for Single variance - Test for equality of two variance for normal population.
Lab Exercises:
1. Test of the single sample mean for known and unknown σ, Test of equality of two means when known and unknown σ.
2. Tests of single variance and equality of variance for large samples
3. Tests for single proportion and equality of two proportion for large samples.

UNIT 3: TESTING OF HYPOTHESIS II Teaching Hours: 15

Tests for single proportion, Tests of equality of two proportions for the normal population, Chi square test for independence of attributes, Chi square test for goodness of fit, Chi square tests for attributes. Concept of confidence interval and confidence coefficient, Confidence intervals for the parameters of univariate normal,

Lab Exercises:

1. Chi-square test for independence of attributes and goodness of fit.
2. To find the confidence interval for different cases of parent normal distribution.

UNIT 4: ANALYSIS OF VARIANCE Teaching Hours: 15

Meaning and assumptions - Fixed, random and mixed effect models - Analysis of variance of one-way and two-way classified data with and without interaction effects, Multiple comparison tests: Tukey's method, critical difference.

Lab Exercises:

1. Construction of one-way and two -way ANOVA
2. Multiple comparison test using Tukey's method and critical difference methods

UNIT 5 NONPARAMETRIC TESTS Teaching Hours: 15

Concept of Nonparametric tests - Run test for randomness - Sign test and Wilcoxon Signed Rank Test for one and paired samples - Run test - Median test and Mann-Whitney-Wilcoxon tests for two samples.

Lab Exercises:

1. Test of one sample and two sample using Run and sign tests.
2. Test of two samples using Run test and Median test.

Essential References

1. Rohatgi V.K and Saleh E, An Introduction to Probability and Statistics, 3rd edition, John Wiley & Sons Inc, New Jersey, 2015.
2. Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 12^{th} edition, Sulthan and Sons, New Delhi, 2020.
3. Kale, B.K. & Muralidharan, K., Parametric Inference: An Introduction, Alpha Science International Ltd., 2015.

Recommended References

1. Rajagopalan M and Dhanavanthan P, Statistical Inference, PHI Learning, New Delhi, 2012.
2. Montgomery, D. C., & Runger, G. C. (2010). Applied statistics and probability for engineers. John wiley & sons.
3. Rajagopalan M and Dhanavanthan P, Statistical Inference, PHI Learning, New Delhi, 2012.

MDS273: FULL STACK WEB DEVELOPMENT

Total Teaching Hours for Semester: 75
No of hours per week: 7(4+3)
Max Marks: 100 **Credits: 4**
Course Type: Major
Course Description
On completion of this course, a student will be familiar with full stack and able to develop a web application using advanced technologies and cultivate good web programming style and discipline by solving the real world scenarios.
Course Outcomes: Upon completion of the course students will be able to

No.	Course Outcomes
CO1	Apply JavaScript, HTML5, and CSS3 effectively to create interactive and dynamic websites
CO2	Describe the main technologies and methods currently used in creating advanced web applications
CO3	Design websites using appropriate security principles, focusing specifically on the vulnerabilities inherent in common web implementations
CO4	Create modern web applications using MEAN

UNIT 1: OVERVIEW OF WEB TECHNOLOGIES AND HTML5 Teaching Hours: 15
Internet and web Technologies- Client/Server model - Web Search Engine-Web Crawling-Web Indexing Search Engine Optimization and Limitations-Web Services –Collective Intelligence –Mobile Web – Features of Web 3.0-HTML vs HTML5-Exploring Editors and Browsers Supported by HTML5-New Elements-HTML5 Semantics-Canvas-HTML Media
Lab Exercises
1. Develop static pages for a given scenario using HTML
2. Creating Web Animation with audio using HTML5 & CSS3
3. Demonstrate Geolocation and Canvas using HTML5

UNIT 2 CLIENT SIDE SCRIPTING Teaching Hours: 15
JavaScript Implementation - Use Javascript to interact with some of the new HTML5 apis -Create and modify Javascript objects- JS Forms - Events and Event handling-JS Navigator-JS Cookies-Introduction to JSON-JSON vs XML-JSON Objects-Importance of Angular JS in web-Angular Expression and Directives-Single Page Application
Lab Exercises
1. Write a JavaScript program to demonstrate Form Validation and Even Handling
2. Create a web application using AngularJS with Forms
3. Implement web application using AJAX with JSON

UNIT 3 XML AND AJAX Teaching Hours: 15
XML-Documents and Vo
cabularies-Versions and Declaration -Namespaces JavaScript and XML: Ajax DOM based

XML processing Event-Transforming XML Documents-Selecting XML Data:XPATH Template based Transformations: XSLT-Displaying XML Documents in Browsers - Evolution of AJAX - Web applications with AJAX -AJAX Framework

Lab Exercises

1. Write an XML file and validate the file using XSD
2. Demonstrate XSL with XSD
3. Demonstrate DOM parser

UNIT 4 SERVER SIDE SCRIPTING Teaching Hours: 15

Introduction to Node.js-REPL Terminal-Package Manager(NPM)-Node.js Modules and file system Node.js Events-Debugging Node JS Application-File System and streams-Testing Node JS with jasmine. **NODE JS WITH MYSQL** Introduction to MySQL- Performing basic database operation (DML) (Insert, Delete, Update, Select).

Lab Exercises

1. Implement a single page web application using AngularJS CRUD Operation using AngularJS
2. Demonstrate Node.js file system module
3. Design a web page to demonstrate CRUD operation using MySQL.

UNIT 4 PYTHON WITH MYSQL Teaching Hours: 15

Installing MySQL Connector for Python. Connecting to MySQL database. Executing SQL queries from Python. Fetching and processing results. Error handling. Reading and writing data to/from MySQL. Data manipulation (e.g., sorting, filtering). Data visualization (e.g., using matplotlib or seaborn)

Lab Exercises

1. Python program to modify your connection code to handle any errors that may occur during the connection process.
2. Building a simple application that uses Python and MySQL (e.g., a CRUD application, data analysis tool)

Essential Reading

1. Internet and World Wide Web:How to Program, Paul Deitel , Harvey Deitel & Abbey Deitel, Pearson Education, 5th Edition, 2018.
2. HTML 5 Black Book (Covers CSS3, JavaScript, XML, XHTML, AJAX, PHP, jQuery), DT Editorial Services, Dreamtech Press, 2nd Edition, 2016.
3. Professional JavaScript for Web Developers, 4th Edition - Matt Frisbie
4. The Node Beginner Book by Manuel Kiessling:A Comprehensive Node.js tutorial.
5. MySQL for Python by Albert Lukaszewski: (2010)

Recommended Reading

1. Chris Northwood, The Full Stack Developer: Your Essential Guide to the Everyday Skills Expected of a Modern Full Stack Web Developer, Apress Publications, 1st Edition, 2018.
2. Laura Lemay, Rafe Colburn & Jennifer Kyrnin, Mastering HTML, CSS & Javascript Web Publishing, BPB Publications, 1st Edition, 2016.
3. Alex Giamas, Mastering MongoDB 3.x, Packt Publishing Limited, First Edition, 2017.

Web Resources:

1. www.w3cschools.com
2. http://www.php.net/docs.php

MDS371: JAVA PROGRAMMING

Total Teaching Hours for Trimester: 75
No of hours per week: 7(4+3)
Max Marks: 100 **Credits: 4**
Course Type: Major

Course Description
This course provides a comprehensive understanding of object-oriented programming structures and principles using JAVA programming. It introduces generics and collections frameworks along with java libraries for implementation of data science applications.

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

No.	Course Outcomes
CO1	Develop an understanding of best practices for writing clean, maintainable, and well-documented code
CO2	Apply object-oriented programming structures in Java to solve problems
CO3	Develop sustainable and innovative solutions for real-time problems.

Cross Cutting Issues:

CO-PO MAPPING:

CO-UNIT MAPPING:
UNIT 1 INTRODUCTION **Teaching Hours:15**
Overview of JVM Introduction to JVM-JVM Architecture-Java Basics- Structure of Java Program- Data Types – Constants –Variables Operators –Conditional statements- Class and Object : Concept - Method Overloading and Overriding - Constructor - this and static keyword

Lab Exercise
1. Implement basic java program
2. Implement the concept of class, data members, member functions and access specifies.
3. Implement the concept of function overloading & Constructor overloading

UNIT 2 ARRAYS AND INHERITANCE **Teaching Hours:15**
Creation and initialization of arrays, one dimensional and multidimensional arrays Inheritance Basics - Multilevel Hierarchy- Using super - Dynamic Method Dispatch -Abstract keyword- Using final with inheritance – Aggregation and Composition in Java
Lab Exercise
1. Implement array processing
2. Implement the concept of inheritance, super, abstract and final keywords.

UNIT 3 INTERFACES AND PACKAGES Teaching Hours:15

Defining Interfaces - Implementing Interfaces - Extending Interfaces- Creating Packages - Importing Packages - Interfaces in a Package. Nested interfaces. Inheritance and interfaces. Use of static in interfaces

Lab Exercise

1. Implement the concept of package
2. Implement the concept of interface

UNIT 4 EXCEPTION HANDLING Teaching Hours:15

Exception Handling in Java-Checked and unchecked exceptions try-catch-finally mechanism - throw statement - throws statement - Built-in-Exceptions – Custom Exceptions-nested try, throw, throws. Introduction to multithreading. – String handling in Java

Lab Exercise

1. Implement the concept of exception Handling
2. Implement string processing

UNIT 5 JAVA I/O Operations Teaching Hours:15

I/O Basics-Streams-Byte Streams-Input Stream classes-Output Stream Classes-Character Streams-Reader Stream classes. File handling in Java. Overview of Collections framework - Introduction to Data Science Libraries.

Lab Exercise

1. Implement File handling
2. Implement the concept of a collection framework

Essential Reading

1. Horstmann, C. S. (2019) Core Java (TM) Volume 1: Fundamentals. Pearson Education India.
2. Richard M.Reese ,Jennifer L Reese ,Alexey Grigorev Java:Data Science made EasyPackt,2017.

Recommended Reading

2. Bloch, J. (2016). Effective java. Pearson Education India.
3. Schildt, H., & Coward, D. (2014). Java: the complete reference. New York: McGraw-Hill Education.

MDS372 - MACHINE LEARNING

Total Teaching Hours for Trimester: 75
No of hours per week: 7 (4+3)
Max Marks: 100 **Credits: 4**
Course Type: Major

Course Description

The objective of this course is to introduce the principles and design of machine learning algorithms. The course is aimed at providing foundations for conceptual aspects of machine learning algorithms along with their applications to solve real world problems.

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

No.	Course Outcomes
CO1	Understand the basic principles of machine learning techniques.
CO2	Understand how machine learning problems are formulated and solved
CO3	Apply machine learning algorithms to solve real world problems.

UNIT 1 INTRODUCTION **Teaching Hours:15**

Machine Learning-Examples of Machine Applications- Learning Associations-Classification-Regression- Unsupervised Learning-Reinforcement Learning. Supervised Learning: Learning class from examples- Probably Approach Correct (PAC) Learning-Noise- Learning Multiple classes. Regression-Model Selection and Generalization.

Lab Exercise

1. Data Exploration using parametric methods
2. Regression analysis

UNIT 2 DIMENSIONALITY REDUCTION **Teaching Hours:15**

Dimensionality Reduction, Dimensionality Reduction: Introduction- Subset Selection-Principal Component Analysis, Feature Embedding-Factor Analysis-Singular Value Decomposition -Multidimensional Scaling

Lab Exercise

3. Data reduction using Principal Component Analysis
4. Data reduction using multi-dimensional scaling

UNIT 3 SUPERVISED LEARNING - I **Teaching Hours:15**

Linear Discrimination: Introduction- Generalizing the Linear Model-Geometry of the Linear Discriminant- Pairwise Separation-Gradient Descent-Logistic Discrimination.
Kernel Machines: Introduction- optical separating hyperplane- v-SVM, kernel tricks- vertical kernel-vertical kernel- defining kernel- multiclass kernel machines- one-class kernel machines.

Lab Exercise

5. Linear discrimination
6. Logistic discrimination
7. Classification using kernel machines.

UNIT 4 SUPERVISED LEARNING - II **Teaching Hours:15**

Multilayer Perceptron: Introduction, training a perceptron- learning Boolean functions- multilayer perceptron- backpropogation algorithm- training procedures.
Combining Multiple Learners : Rationale-Generating diverse learners- Model combination schemes-voting, Bagging- Boosting- fine tuning an Ensemble.

Lab Exercise

8. Classification using MLP
9. Ensemble Learning

UNIT 5 UNSUPERVISED LEARNING **Teaching Hours:15**

Clustering Introduction-Mixture Densities, K-Means Clustering- Expectation-Maximization algorithm- Mixtures of Latent Varaible Models-Supervised Learning after Clustering - Hierachial Clustering-Clustering- Choosing the number of Clusters.

Lab Exercise

10. K means clustering
11. Hierarchical clustering

Essential Reading

1. E. Alpaydin, Introduction to Machine Learning, 3rd Edition, MIT Press, 2014.

Recommended Reading

1. C.M.Bishop,PatternRecognitionandMachineLearning,Springer,2016.
2. T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, 2nd Edition,2009
3. K.P.Murphy,MachineLearning:AProbabilisticPerspective,MITPress,2012.

MDS331: REGRESSION MODELING

Total Teaching Hours for Trimester: 45
No of hours per week: 4
Max Marks: 100 **Credits: 3**
Course Type: Major

Course Description
This course deals with linear and non-linear regression models with their assumptions, estimation and test of significance of regression coefficients, and overall regression model with various model selection criteria.
Course Outcomes: Upon completion of the course students will be able to

No.	Course Outcomes
CO1	Formulate the linear regression model and its application to real data.
CO2	Understand and identify the various assumptions of linear regression models.
CO3	Identify the correct model using model selection and variable selection criteria.
CO4	Ability to use and understand generalizations of the linear model to binary and count data.

Unit 1 **Teaching Hours: 10**

Simple Linear Regression

Introduction to regression analysis: overview and applications of regression modelling, major steps in regression modelling. Simple linear regression: assumptions, estimation of regression coefficients using ordinary least squares and maximum likelihood estimation, properties of regression coefficients, significance and confidence intervals of regression coefficients.

Unit 2 **Teaching Hours: 9**

Multiple Linear Regression

Assumptions, ordinary least square estimation of regression coefficients, properties of the regression coefficients, significance and confidence intervals of regression coefficients with interpretation.

Unit 3 **Teaching Hours: 9**

Model Adequacy

Residual analysis; Departures from underlying assumptions: Multicollinearity, Heteroscedasticity, Autocorrelation, Effect of outliers. Diagnostics and remedies.

Unit 4 **Teaching Hours: 8**

Model Selection Criteria

Model selection criteria: R-Square, Adjusted R- Square, Mean Square error criteria; Variable selection criteria: Forward, Backward and Stepwise procedures.

Unit 5 **Teaching Hours: 9**

Non-Linear Regression

Introduction to nonlinear regression, Least squares in the nonlinear case and estimation of parameters, Models for binary and count response variable.

Essential Reading

1. Montgomery D.C, Peck E.A and Vining G.G, Introduction to Linear Regression Analysis, John Wiley and Sons Inc,. New York, 2012.
2. Chatterjee S and Hadi A, Regression Analysis by Example, 4th edition, John Wiley and Sons Inc, New York, 2015.

Recommended Reading

1. George A.F.S and Lee A.J, Linear Regression Analysis, John Wiley and Sons, Inc, 2012.
2. Pardoe I, Applied Regression Modeling, John Wiley and Sons Inc, New York, 2012
3. Iain Pardoe, Applied Regression Modeling, John Wiley and Sons, Inc, 2012.
4. P. McCullagh, J.A. Nelder, Generalized Linear Models, Chapman & Hall, 1989.

MDS341A: CATEGORICAL DATA ANALYSIS

Total Teaching Hours for Semester: 45

No of hours per week: 4

Max Marks: 100 **Credits: 3**

Course Type: Elective Course Description

Categorical data analysis deals with the study of information captured through expressions or verbal forms. This course equips the students with the theory and methods to analyse and categorical responses.

Course Outcomes

No.	Course Outcomes
CO1	Describe the categorical response
CO2	Identify tests for contingency tables
CO3	Apply regression models for categorical response variables
CO4	Analyse contingency tables using log-linear models

UNIT 1 **Teaching Hours:9**
Introduction

Categorical response data - Probability distributions for categorical data - Statistical inference for discrete data

UNIT 2 **Teaching Hours:9**
Contingency Tables

Probability structure for contingency tables - Comparing proportions with 2x2 tables - The odds ratio - Tests for independence - Exact inference

UNIT 3 **Teaching Hours:9**
Generalised Linear Model

Components of a generalised linear model - GLM for binary and count data - Statistical inference and model checking - Fitting GLMs

UNIT 4 **Teaching Hours:9**
Logistic Regression

Interpreting the logistic regression model - Inference for logistic regression - Logistic regression with categorical predictors - Multiple logistic regression - Summarising effects - Building and applying logistic regression models

UNIT 5 **Teaching Hours:9**
Log-linear models for Contingency Tables

Loglinear models for two-way and three-way tables - Inference for Loglinear models - the log-linear-logistic connection - Independence graphs and collapsibility – Models for matched pairs: Comparing dependent proportions.

Essential Reading

[1] Agresti, A. (2012). *Categorical Data Analysis*, 3rd edition. New York, Wiley

Recommended References

[1] Le, *C.T. (2009). Applied Categorical Data Analysis and Translational Research*, 2nd edition, John Wiley and Sons.
[2] Agresti, A. (2010). *Analysis of ordinal categorical*. John Wiley & Sons.
[3] Stokes, M. E., Davis, C. S., & Koch, G. G. (2012). *Categorical data analysis using SAS*. SAS Institute.
[4] Agresti, A. (2018). *An introduction to categorical data analysis*. John Wiley & Sons.
[5] Bilder, C. R., & Loughin, T. M. (2014). *Analysis of categorical data with R*. Chapman and Hall/CRC.

MDS341B: MULTIVARIATE ANALYSIS

Total Teaching Hours for Semester: 45

No of hours per week: 4
Max Marks: 100 **Credits: 3**

Course Type: Elective Course Description
This course lays the foundation of Multivariate data analysis. The exposure provided to the multivariate data structure, multinomial and multivariate normal distribution, estimation and testing of parameters, and various data reduction methods would help the students in having a better understanding of research data, its presentation, and analysis.

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

No.	Course Outcomes
CO1	Understand multivariate data structure, multinomial, and multivariate normal distribution.
CO2	Apply likelihood Ratio tests for multivariate normal proportions
CO3	Analyze multivariate data using (MANOVA) of one and two-way classified data.

UNIT 1 **Teaching Hours:9**
Introduction
Basic concepts on the multivariate variable. Bivariate normal distribution; an overview. Multivariate normal distribution and its properties, Its expectation, and Variance-Covariance matrix. Conditional distributions and Independence of random vectors. Multinomial distribution.

UNIT 2 **Teaching Hours:9**
Distribution

Sample mean vector and its distribution. Likelihood ratio tests: Tests of hypotheses about the mean vectors and covariance matrices for multivariate normal populations.

UNIT 3 **Teaching Hours:9**

Multivariate Analysis

Multivariate analysis of variance (MANOVA) of one and two- way classified data. Multivariate analysis of covariance. Wishart distribution, Hotelling's T2 and Mahalanobis' D^2 statistics and their properties.

UNIT 4 **Teaching Hours:9**

Classification and Discriminant Procedures Bayes, minimax, and Fisher's criteria for discrimination between two multivariate normal populations. Sample discriminant function. Tests associated with discriminant functions. Probabilities of misclassification and their estimation.

UNIT 5 **Teaching Hours:9**

Principal Component and Factor Analysis

Principal components, sample principal components asymptotic properties. Canonical variables and canonical correlations: definition, estimation, computations. Factor analysis: Orthogonal factor model, factor loadings, estimation of factor loadings.

Essential Reading

[1] Anderson, T.W. 2009. An Introduction to Multivariate Statistical Analysis, 3rd Edition, John Wiley.
[2] Everitt B, Hothorn T, 2011. An Introduction to Applied Multivariate Analysis with R, Springer.
[3] Barry J. Babin, Hair, Rolph E Anderson, and William C. Blac, 2013, Multivariate Data Analysis, Pearson New International Edition.

Recommended Reading

[1] Giri, N.C. 1977. Multivariate Statistical Inference. Academic Press.
[2] Chatfield, C. and Collins, A.J. 1982. Introduction to Multivariate analysis. Prentice Hall.
[3] Srivastava, M.S. and Khatri, C.G. 1979. An Introduction to Multivariate Statistics. North- Holland.

MDS342C: STOCHASTIC PROCESSES

Total Teaching Hours for Semester: 45
No of hours per week: 4L-0-0P
Max Marks: 100 **Credits: 3**

Course Type: Discipline Specific Elective Course Description

This course is designed to introduce the concepts of theory of estimation and testing of hypotheses. This paper also deals with the concept of parametric tests for large and small samples. It also provides knowledge about non-parametric tests and its applications.

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

No.	Course Outcomes

CO1	Understand and apply the types of stochastic processes in various real-life scenarios.
CO2	Demonstrate a discrete space stochastic process in a discrete index and estimate the evolving time in a state.
CO3	Apply probability arguments to model and estimate the counts in continuous time.
CO4	Evaluate the extinction probabilities of a generation.
CO5	Development of renewal equations in discrete and continuous time.
CO6	Understand the stationary process and application in Time Series Modelling

UNIT 1 Teaching Hours: 09

INTRODUCTION TO STOCHASTIC PROCESSES

Classification of Stochastic Processes, Markov Processes – Markov Chain - Countable State Markov Chain. Transition Probabilities, Chapman - Kolmogorov's Equations, Calculation of n - step Transition Probability and its limit.

UNIT 2 Teaching Hours: 09

POISSON PROCESS

Classification of States, Recurrent and Transient States - Transient Markov Chain, Random Walk. Continuous Time Markov Process: Poisson Processes, Birth and Death Processes, Kolmogorov's Differential Equations, Applications.

UNIT 3 Teaching Hours: 09

BRANCHING PROCESS

Branching Processes – Galton – Watson Branching Process - Properties of Generating Functions – Extinction Probabilities – Distribution of Total Number of Progeny.

UNIT 4 Teaching Hours: 09

RENEWAL PROCESS

Renewal Processes – Renewal Process in Discrete and Continuous Time – Renewal Interval –Renewal Function and Renewal Density – Renewal Equation – Renewal theorems: Elementary Renewal Theorem.

UNIT 5 Teaching Hours: 09

STATIONARY PROCESS

Stationary Processes: Application to Time Series. Auto-covariance and Auto-correlation functions and their properties. Moving Average, Autoregressive, Autoregressive Moving

Average. Basic ideas of residual analysis, diagnostic checking, forecasting.

Essential References

[1] Stochastic Processes, R.G Gallager, Cambridge University Press, 2013.

[2] Stochastic Processes, S.M Ross, Wiley India Pvt. Ltd, 2008.

Recommended References

[1] Stochastic Processes from Applications to Theory, P.D Moral and S. Penev, CRC Press, 2016.

[2] Introduction to Probability and Stochastic Processes with Applications, B..C. Liliana, A Viswanathan, S. Dharmaraja, Wiley Pvt. Ltd, 2012.

MDS311: CLOUD ESSENTIALS

Total Teaching Hours for Semester: 30

No of hours per week:2

Max Marks: 100 **Credits: 2**

Course Type: Major Course Description

This on-line course gives students an overview of the field of Cloud Computing, its enabling technologies, main building blocks, and hands-on experience through projects utilizing public cloud infrastructures (Amazon Web Services (AWS) and Microsoft Azure). The student learns the topics of cloud infrastructures, virtualization, software defined networks and storage, cloud storage, and programming models.

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

No.	Course Outcomes
CO1	Understand the *core concepts* of the cloud computing paradigm.
CO2	Apply fundamental concepts of cloud *infrastructures, cloud storage* and in storage systems such as Amazon S3 and HDFS.
CO3	Analyze various *cloud programming models* and apply them to solve problems on the cloud.

UNIT 1 **Teaching Hours:6**

Introduction:
Definition and evolution of Cloud Computing, Enabling Technologies, Service and Deployment Models Popular Cloud Stacks and Use Cases Benefits, Risks, and Challenges of Cloud Computing Economic Models and SLAs

UNIT 2 **Teaching Hours:6**

Cloud Infrastructure:

Historical Perspective of Data Centers, Datacenter Components: IT Equipment and Facilities

Design Considerations: Requirements, Power, Efficiency, & Redundancy, Power Calculations, PUE and Challenges in Cloud Data Centers, Cloud Management and Cloud Software Deployment Considerations.

UNIT 3 **Teaching Hours:6**

Virtualization:

Virtualization (CPU, Memory, I/O),Case Study: Amazon EC2,Software Defined Networks (SDN),Software Defined Storage (SDS)

UNIT 4 **Teaching Hours:6**

Cloud Storage:

Introduction to Storage Systems, Cloud Storage Concepts, Distributed File Systems (HDFS, Ceph FS) Cloud Databases (HBase, MongoDB, Cassandra, DynamoDB) ,Cloud Object Storage (Amazon S3, OpenStack Swift, Ceph)

UNIT 5 **Teaching Hours:6**

Programming Models:

Distributed Programming for the Cloud Data-Parallel Analytics with Hadoop MapReduce (YARN)

Essential Reading:

[1] Douglas Corner The Cloud Computing Book: The Future of Computing Explained,CRC Press,2021

[2] Chellammal Surianarayanan,Essentials of Cloud Computing: A Holistic Perspective, Springer, 2019.

Recommended Reading:

[1] K. Chandrasekaran,Essentials of Cloud Computing,CRC press,2014

MDS431: Data Driven Modelling and Visualization

Total Teaching Hours for Trimester: 30
No of hours per week: 3
Max Marks: 100 **Credits: 2**
Course Type: Major
Course Description: This course provides an overview of how to analyse, interpret, and communicate insights from data. A combination of lectures, hands-on exercises, and real-world projects, students will learn how to leverage data effectively to build statistical and machine learning models, uncover patterns.

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

No.	Course Outcomes
CO1	Analyse data to identify trends, patterns, outliers
CO2	Evaluate Charts which could present the insight effectively
CO3	Present Data Insights using Charts and Dashboards

Unit 1 **Teaching Hours: 10**
Data Processing and Analysis
Collection of relevant data from Structured (spreadsheet, database, data warehouse) and Unstructured Data (text, images, sensors). Filtering, Aggregation, Grouping, Pivoting, Scaling, Managing Data Type. Data Profiling, Summarizing the main characteristics of the data, identifying patterns, trends, and outliers. Visualization techniques for data analysis: trend lines, histograms, scatter plots, box plots.
Unit 2 **Teaching Hours: 6**
Data Visualisation
Overview of Visualisation Tools. Understand visualization need. Comparison, Composition, Relation, Distribution. Use of Statistical Functions, searching for insights using Query or Pivot. Insights from Pie Charts, Area Charts, TreeMaps, Correlation Charts, Donut Charts
Case Study: Google Trends, Tableau Public Gallery.
Unit 3 **Teaching Hours: 6**
Advanced Data Visualisation
Dashboard, common filters across charts, Animation, Storytelling with Data Visualisations
Stock Chart, Candlestick Chart, Sunburst Diagram, Word Clouds, Waterfall chart, Funnel Chart, Polar Graph, GeoSpatial Map, Gantt Chart, Choropleth Map, Parallel Coordinates Plot, Non-Ribbon Chord Diagram.
Case Study: MakeOverMonday: A Social Data Project

Essential References

1. O'Connor, Errin. Microsoft Power BI Dashboards Step by Step. Microsoft Press, 2018.
2. Milligan, Joshua N, Learning Tableau, Packt Publishing Ltd, 2019

Recommended Reading

1. Tufte, E., “The Visual display of quantitative information”, Second Edition, 2002

2. Few, Stephen, “Information Dashboard Design.”, 2013
3. Knaflic, Cole Nussbaumer, “Storytelling with Data: Let's Practice!”. John Wiley & Sons,2019

MDS432 Time Series Analysis and Forecasting Techniques

Total Teaching Hours for Trimester: 60
No of hours per week: 5
Max Marks: 100 **Credits: 4**
Course Type: Core

Course Description
This course covers applied statistical methods pertaining to time series and forecasting techniques. Moving average models like simple, weighted and exponential are dealt with. Stationary time series models and non-stationary time series models like AR, MA, ARMA and ARIMA are introduced to analyse time series data.

Course Outcomes

No.	Course Outcomes
CO1	Ability to approach and analyse univariate time series
CO2	Ability to differentiate between various time series models like AR, MA, ARMA and ARIMA models
CO3	Evaluate stationary and non-stationary time series models
CO4	Able to forecast future observations of the time series

Unit 1 **Teaching Hours: 12**

INTRODUCTION TO TIME SERIES
Introduction to time series and stochastic process, graphical representation, components and classical decomposition of time series data. Auto-covariance and auto-correlation functions, Exploratory time series analysis, Test for trend and seasonality, Smoothing techniques such as Exponential and moving average smoothing.

Unit 2 **Teaching Hours: 12**
STATIONARY TIME SERIES MODELS
Wold representation of linear stationary processes, generalised linear models, Study of linear time series models: Autoregressive, Moving Average and Autoregressive Moving average models and their statistical properties like ACF and PACF function.

Unit 3 **Teaching Hours: 12**
ESTIMATION OF ARMA MODELS
Estimation of ARMA models: Yule- Walker estimation of AR Processes, Maximum likelihood and least squares estimation for ARMA Processes, MMSE forecast and l-step ahead forecast, Residual analysis and diagnostic checking.

Unit 4 **Teaching Hours: 12**
NON-STATIONARY TIME SERIES MODELS
Concept of non-stationarity, general unit root tests for testing non stationarity; basic formulation of the ARIMA Model and their statistical properties-ACF and PACF; forecasting

using ARIMA models.

Unit 5 **Teaching Hours: 12**

INTRODUCTION TO MULTIVARIATE AND SEASONAL TIME SERIES MODELS

Stationary Multivariate Time series, Vector AR models, Vector MA models, Vector ARMA models- its stationarity properties, Non stationarity and Cointegration.

Seasonal time series models, Introduction to SARIMA models, different representations of SARIMA models and its forecast.

Essential References

1. George E. P. Box, G.M. Jenkins, G.C. Reinsel and G. M. Ljung, Time Series
2. analysis Forecasting and Control, 5th Edition, John Wiley & Sons, Inc., New Jersey, 2016.
3. Montgomery D.C, Jennigs C. L and Kulachi M,Introduction to Time Series
4. analysis and Forecasting, 2nd Edition,John Wiley & Sons, Inc., New Jersey, 2016.

Recommended Reading

1. Brockwell, P.J and Davis R.A. (2006) Time Series: Theory and Methods, 2^{nd} edition, Spinger.
2. Shumway, R. H and Stoffer, D. S. (2006). Time series Analysis and its Applications. Springer.
3. Anderson T.W,*Statistical Analysis of Time Series*, John Wiley&Sons, Inc., New Jersey,1971.

MDS471-NEURAL NETWORKS AND DEEP LEARNING

Total Teaching Hours for Trimester: 75

No of hours per week: 7 (4+3)

Max Marks: 100 **Credits: 4**

Course Type: Major

Course Description

The main aim of this course is to provide fundamental knowledge of neural networks and deep learning and its implementation. On successful completion of the course, students will acquire fundamental knowledge of neural networks and deep learning, such as Basics of neural networks, shallow neural networks, deep neural networks, forward & backward propagation process and build various research projects.

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

No.	Course Outcomes
CO1	Understand the fundamental concepts of Artificial Neural Networks (ANN) and their evolution, Analyze the theory and architecture of shallow neural networks, implementing learning factors in Back-Propagation Networks for effective training.
CO2	Apply convolutional operations for image recognition in Convolutional Neural Networks (CNN) and implement different CNN architectures.
CO3	Evaluate the challenges in training Recurrent Neural Networks (RNN) and create an implementation of Long Short-Term Memory (LSTM) for sequential data analysis. Understand and apply features of Auto encoders and Restricted Boltzmann Machines (RBM) for efficient unsupervised feature learning. Apply Neural network models to solve real time problems.

Unit 1 **Teaching Hours: 15**

INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS

Neural Networks-Application Scope of Neural Networks- Fundamental Concept of ANN: The Artificial Neural Network-Biological Neural Network-Comparison between Biological Neuron and Artificial Neuron-Evolution of Neural Network. Basic models of ANN-Learning Methods-Activation Functions-Importance Terminologies of ANN.

Lab Exercise

1. Create a program to build and train an Artificial Neural Network
2. Create a regression model with Artificial Neural Networks (ANN) by following steps like data preparation and model training for predicting continuous outcomes

Unit 2 **Teaching Hours: 15**

SUPERVISED LEARNING NETWORK

Shallow neural networks- Perceptron Networks-Theory-Perceptron Learning Rule Architecture- Perceptron Training Algorithm for Single and Multiple Output Classes.

Back Propagation Network- Theory-Architecture-Training Algorithm-Learning Factors for Back-Propagation Network.

Radial Basis Function Network RBFN: Theory, Architecture and Algorithm.

Lab Exercise

1. Develop a shallow neural network emphasizing simplicity and clarity in steps such as data processing, building the network architecture, and training.
2. Create a program to implement learning factors in Backpropagation Neural Networks (BPN), focusing on steps such as data handling, network architecture, and the incorporation of learning factors for improved training.

Unit 3 **Teaching Hours: 15**

CONVOLUTIONAL NEURAL NETWORK

Introduction - Components of CNN Architecture - Rectified Linear Unit (ReLU) Layer - Exponential Linear Unit (ELU, or SELU), types of CNN Architectures, alexnet, zfnet, googlenet and VGG -Applications of CNN.

Lab Exercise

1. Ilustrate the step-by-step process of applying convolutional operations on data for enhanced understanding
2. Develop a Convolutional Neural Network (CNN), guiding through the stages of data preprocessing, model design, and training for effective image recognition

Unit 4 **Teaching Hours: 15**

RECURRENT NEURAL NETWORK

Introduction- The Architecture of Recurrent Neural Network- The Challenges of Training Recurrent Networks - Long Short-Term Memory (LSTM) - Applications of RNN.

Lab Exercise

1. Construct a Recurrent Neural Network (RNN) it includes key steps such as data preprocessing, model architecture design, and training to capture sequential dependencies in data
2. Create a Long Short-Term Memory (LSTM) implementation guiding through essential steps such as data preparation, designing the LSTM model architecture, and training for effective sequential data analysis.

Unit 5 **Teaching Hours: 15**

AUTO ENCODER AND RESTRICTED BOLTZMANN MACHINE

Introduction - Features of Auto encoder Types of Auto encoder Restricted Boltzmann Machine- Boltzmann Machine - RBM Architecture -Example - Types of RBM.

Lab Exercise

1. Develop a Restricted Boltzmann Machine (RBM) implementation using the essential steps of data handling, model construction, and training for unsupervised learning tasks.
2. Build an Autoencoder for efficient unsupervised feature learning.

Essential Reading

1. S.N.Sivanandam, S. N. Deepa, Principles of Soft Computing, Wiley-India, 3rd Edition, 2018.
2. Deep Learning Ian Goodfellow Yoshua Bengio Aaron Courville, MIT PRESS, Ist Edition,2020

Recommended Reading

1. Charu C. Aggarwal, Neural Networks and Deep Learning, Springer, July 2023.
2. Francois Chollet, Deep Learning with Python, Manning Publications; second edition, 2021

3. John D. Kelleher, Deep Learning (MIT Press Essential Knowledge series), The MIT Press, 2019.

MDS472A: WEB ANALYTICS

Total Teaching Hours for Trimester: 60
No of hours per week: 6 (3+3)
Max Marks: 100 **Credits: 3**
Course Type: Elective
Course Description
The objective of this course is to provide overview and importance of Web analytics in terms of visualizations. This course also explores the effective of Web analytic strategies and implementation using Google analytics with visual analytics.

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

No.	Course Outcomes
CO1	Understand the concept and importance of Web analytics in an organization and the role of Web analytic in collecting, analyzing and reporting website traffic.
CO2	Identify key tools and diagnostics associated with Web analytics.
CO3	Explore effective Web analytics strategies and implementation and Understand the importance of web analytic as a tool for e-Commerce, business research, and market research.

Unit 1 **Teaching Hours: 12**
INTRODUCTION TO WEB ANALYTICS
Introduction to Web Analytics: Web Analytics Approach – A Model of Analysis – Context matters – Data Contradiction – Working of Web Analytics: Log file analysis – Page tagging – Metrics and Dimensions – Interacting with data in Google Analytics
Lab Exercise
1. Working concept of web analytics
   Create & Hosting of Website, Creation of Google Analytics
   Account and Adding Property, Container Creation in Google Tag Manager, Create A New Tag and tag the Website Created.
2. Measuring Enhance Metrics

Unit 2 **Teaching Hours: 12**
LEARNING ABOUT USERS THROUGH WEB ANALYTICS
Goals: Introduction – Goals and Conversions – Conversion Rate – Goal reports in Google Analytics – Performance Indicators – Analyzing Web Users: Learning about users – Traffic Analysis – Analyzing user content – Click-Path analysis – Segmentation
Lab Exercise
1. Web Analytics: Log file analysis
2. Explore all the available Metrics for Google Demo Account Data with and without Filter.

Unit 3 **Teaching Hours: 12**
WORKING WITH ANALYTICS
Different analytical tools - Key features and capabilities of Google analytics- How Google analytics works - Implementing Google analytics - Getting up and running with Google analytics -Navigating Google analytics – Using Google analytics reports -Google metrics - Using visitor data to drive website improvement- Focusing on key performance indicators-

Integrating Google analytics with third-Party applications.

Lab Exercise

1. Create an Event by using all Built-In Variables of Pages and Scrolling.
2. Create youtube video view event by using all the matching Conditions like equal, contains etc in the trigger.

Unit 4 **Teaching Hours: 12**

OVERVIEW OF QUALITATIVE ANALYSIS

Lab Usability Testing- Heuristic Evaluations- Site Visits- Surveys (Questionnaires) - Testing and Experimentation: A/B Testing and Multivariate Testing-Competitive Intelligence - Analysis Search Analytics: Performing Internal Site Search Analytics, Search Engine Optimization (SEO) and Pay per Click (PPC)-Website Optimization against KPIs- Content optimization- Funnel/Goal.

Lab Exercise

1. Create a Custom Event to trace the Registered Users using Button Click and enable the conversion rate for the same (with all Built-In Variables of click)
2. Create funnel optimization using Google Analytics.

Unit 5 **Teaching Hours: 12**

VISUAL ANALYTICS

Drill down and hierarchies-Sorting-Grouping- Additional Ways to Group- Creating Sets-Analysis with Cubes and MDX- Filtering for Top and Top N- Using the Filter Shelf- The Formatting Pane- Trend Lines- Forecasting.

Lab Exercise

1. Visualization

Essential Reading

1. Beasley M, Practical web analytics for user experience: How analytics can help you understand your users. Newnes, 1st edition, Morgan Kaufmann, 2013.
2. Sponder M, Social media analytics: Effective tools for building, interpreting, and using metrics, 1st edition, McGraw Hill Professional, 2013.
3. Clifton B, Advanced Web Metrics with Google Analytics, 3rd edition, John Wiley & Sons, 2012.

Recommended Reading

1. Peterson E. T, Web Analytics Demystified: AMarketer's Guide to Understanding How Your Web Site Affects Your Business. Ingram, 2004.
2. Sostre P, LeClaire J, Web Analytics for dummies, John Wiley & Sons,2007.
3. Burby J, Atchison S, Actionable web analytics: using data to make smart business decisions, John Wiley & Sons, 2007.
4. Dykes B, Web analytics action hero: Using analysis to gain insight and optimize your business, Adobe Press, 2011.

MDS472B: IOT ANALYTICS

Total Teaching Hours for Trimester: 60
No of hours per week: 6(3+3)
Max Marks: 100 **Credits: 3**
Course Type: Elective

Course Description
This course offers an opportunity to comprehend the principles of big data analytics within the context of the Internet of Things (IoT). Emphasis is placed on comprehending architectural components, protocols for application development, and selecting data analytics and visualization tools tailored to specific problem domains. Through hands-on experiences, students will engage in data collection, storage, and analysis procedures pertaining to IoT data.

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

No.	Course Outcomes
CO1	Illustrate the process of constructing a data flow for linking IoT system or device data to the cloud utilizing particular formats.
CO2	Describe the utilization of big data tools in distributed computing for processing IoT data
CO3	Employ algorithms to analyze IoT data patterns and extract intelligence

Unit 1 **Teaching Hours: 12**
INTRODUCING IOT ANALYTICS
IoT data and Big data- -IoT Analytics Lifecycle and Techniques- IoT Data Collection, IoT Data Analysis, IoT Data Deployment, Operationalization, and Reuse- Defining IoT analytics and challenges- IoT, Cloud and Big Data Integration for IoT Analytics **-**Cloud-based IoT Platforms- Requirements of IoT Big Data Analytics Platform- Functional Architecture-Data Analytics for the IoT.

Lab Exercise

1. Overview of IoT data sources, formats and data preprocessing techniques for IoT data.
2. Exploratory Data Analysis (EDA) for IoT Data.

Unit 2 **Teaching Hours: 12**
IOT DEVICES AND NETWORKING PROTOCOLS
The Wild World of IoT Devices-Sensor Types-Networking Basics- IoT Networking Connectivity Protocols- IoT Networking Data Messaging Protocols-MQTT- HTTP and IoT-REST- CoAP- Analyzing Data to Infer Protocol and Device.
Lab Exercise

1. Analyzing periodicity and autocorrelation in IoT time series data.
2. Building predictive models for IoT data (e.g., regression, time series forecasting).

Unit 3 **Teaching Hours: 12**

IOT ANALYTICS FOR THE CLOUD

Building Elastic Analytics-Cloud Infrastructure- Elastic Analytics Concepts- Introduction to Building an IoT Analytics Pipeline on Google Cloud, AWS, Azure, ThingSpeak

Lab Exercise

1. Building real-time analytics pipelines for IoT data. Implementing windowing and aggregation operations on streaming IoT data.
2. Preparing IoT Cloud setup-AWS EC2.

Unit 4 Teaching Hours: 12

EXPLORING IOT DATA

Exploring and Visualizing Data-Techniques to understand Data Quality- Data Completeness-Data Validity- Assessing Information Lag-Representativeness- Basic Time Series Analysis-The Basics of Geospatial Analysis.

Lab Exercise

1. Loading sensor data to Cloud.
2. IoT Data Visualization and Dashboarding. Building interactive dashboards for monitoring IoT systems.

Unit 5 Teaching Hours: 12

DATA SCIENCE FOR IOT ANALYTICS

Machine Learning- Representation-Evaluation-Optimization-Generalization-Feature Engineering-Dealing with missing Values-Time Series Handling, Validation Methods-Understanding Bias-Variance Tradeoff- Machine Learning Models- Use cases for Deep Learning with IoT Data- Data Analytics in Smart Buildings.

Lab Exercise

1. Deploying machine learning models for IoT data analysis.
2. Scalable machine learning techniques for IoT big data analytics

Essential Reading

1. Andrew Minteer, Analytics for the Internet of Things(IoT), Packt Publishing, First Edition, 2017.
2. Tausifa Jan Saleem and Mohammad Ahsan Chishti, Big Data Analytics for Internet of Things, Wiley, First Edition, 2021.

Recommended Reading

1. John Soldatos, Building Blocks for IoT Analytics, River Publishers, First Edition, 2017.
2. Harry G. Perros, An Introduction to IoT Analytics, CRC Press, First Edition, 2021.

Web Resources:

1. www.ThingSpeak.com

MDS472C: NATURAL LANGUAGE PROCESSING

Total Teaching Hours for Trimester: 60
No of hours per week: 6(3+3)
Max Marks: 100 **Credits: 3**
Course Type: Elective

Course Description
The course introduces building blocks of Natural Language Processing pipeline. It provides comprehensive understanding on the methods and applications of NLP in the current data analysis paradigms.

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

No.	Course Outcomes
CO1	Understand word and sentence level analysis
CO2	Apply Vector semantics and embeddings for representation of text
CO3	Design text based information retrieval systems
CO4	Analyze NLP applications for real world data

Unit 1 **Teaching Hours: 12**
PARSING AND SYNTAX
Introduction to NLP- Background and overview- NLP Applications -NLP hard Ambiguity-Algorithms and models, Knowledge Bottlenecks in NLP- Introduction to NLTK Word Level Analysis: Regular Expressions, Text Normalization, Edit Distance, Parsing and Syntax-Spelling, Error Detection and correction
Lab Exercise
1. Application of Regular Expression
2. Exploratory Data Analysis for text data

Unit 2 **Teaching Hours: 12**
SEQUENCE LABELING FOR PARTS OF SPEECH AND NAMED ENTITIES
Words and Word classes- English Word Classes, Part-of-Speech Tagging, Named Entities and Named Entity Tagging, HMM Part-of-Speech Tagging
SEMANTIC ANALYSIS AND DISCOURSE PROCESSING
Semantic Analysis: Meaning Representation-Lexical Semantics- Ambiguity-Word Sense Disambiguation. Discourse Processing: cohesion-Reference Resolution- Discourse Coherence and Structure.
Lab Exercise
1. Demonstration of POS and NER
2. Demonstration of word sense disambiguation

Unit 3 **Teaching Hours: 12**
VECTOR SEMANTICS AND EMBEDDINGS
Lexical Semantics, Vector semantics, words and vectors, Cosine for measuring similarity, TF-IDF: Weighing terms in the vector, Pointwise Mutual Information (PMI), Word2vec.
Lab Exercise

1. Demonstration of Vectorization
2. Demonstration of word2Vec

Unit 4 **Teaching Hours: 12**

QUESTION ANSWERING AND INFORMATION RETRIEVAL

Information Retrieval, Document Scoring, Term weighting and document scoring, Inverted Index, Evaluation of Information-Retrieval Systems, Using Neural IR for Question Answering, Evaluating Retrieval-based Question Answering

Lab Exercise

1. Demonstration of information retrieval system using tf-idf.
2. Demonstration of information retrieval system using dense Vectors.

Unit 5 **Teaching Hours: 12**

NLP APPLICATIONS

Language Divergences and Typology, Machine Translation using Encoder-Decoder, Translating in low-resource situations, MT Evaluation, Chatbots & Dialogue Systems, Properties of Human Conversation, Dialogue Acts and Dialogue State, Dialogue Acts and Dialogue State, Training chatbots, Fine Tuning for Quality and Safety, Learning to perform retrieval as part of responding, RLHF

Lab Exercise

1. Implementation of NLP Application

Essential Reading

1. Speech and Language Processing, Daniel Jurafsky and James H., 3rd Edition, Martin Prentice Hall,2023.
2. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press, 1999.

Recommended Reading

1. Foundations of Computational Linguistics: Human-computer Communication in Natural Language, Roland R. Hausser, Springer, 2014.
2. Steven Bird, Ewan Klein and Edward Loper Natural Language Processing with Python and spacy, O'Reilly Media; 1 edition, 2009.

MDS472D: Graph Analytics

Total Teaching Hours for Trimester: 60
No of hours per week: 6(3+3)
Max Marks: 100 **Credits: 3**
Course Type: Elective

Course Description
The course aims to equip students with a comprehensive understanding of graph theory, algorithms, and their applications in data science. Students will explore fundamental concepts of graph analytics, learn various graph algorithms, develop practical skills in analyzing graph data, understand advanced topics such as community detection and graph-based machine learning, and apply graph analytics techniques to real-world datasets and problems.

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

No.	Course Outcomes
CO1	Understanding of Graph Theory Fundamentals: Students will demonstrate a solid understanding of fundamental concepts in graph theory.
CO2	Proficiency in Graph Algorithms: Students will be proficient in implementing and applying common graph algorithms.
CO3	Application of Community Detection Techniques: Students will be able to apply community detection algorithms.
CO4	Knowledge of Graph-Based Machine Learning: Students will gain knowledge of graph-based machine learning techniques and understand their applications.
CO5	Practical Application of Graph Analytics: Students will apply graph analytics techniques to real-world datasets and problems.

Unit 1 **Teaching Hours: 12**
Introduction to Graphs and Graph Theory

Types of graphs: directed, undirected, weighted, etc. Graph representations: adjacency matrix, adjacency list, and edge list, Basic graph properties and terminology, Graph operations and transformations.

Lab Exercise

1. Graph Construction and Manipulation: Load a small real-world dataset (social network, citation network) and construct different representations. Implement basic transformations.
2. Graph Visualization: Using libraries like NetworkX, visualize different graph types highlighting their properties.

Unit 2 **Teaching Hours: 12**
Graph Algorithms and Centrality Measures

Breadth-first search (BFS) and depth-first search (DFS), Shortest path algorithms: Dijkstra's algorithm, Bellman-Ford algorithm, Centrality measures: degree centrality, betweenness centrality, closeness centrality, PageRank algorithm and its applications, Applications of graph algorithms in social networks and recommendation systems.

Lab Exercise

1. Traversal and Pathfinding: Implement BFS and DFS. Apply Dijkstra's algorithm to find shortest paths on a transportation or routing dataset.
2. Evaluating Influence: Calculate centrality measures on a social network dataset. Analyze results to identify important nodes.

Unit 3 **Teaching Hours: 12**

Community Detection and Network Analysis

Introduction to community detection and modularity, Common community detection algorithms: Louvain method, Girvan-Newman algorithm, Network motifs and subgraph analysis, Structural balance theory and triadic closure, Application of community detection in social network analysis and biological networks

Lab Exercise

1. Community Finding: Apply Louvain and Girvan-Newman algorithms on a dataset with known community structure (or a larger social dataset). Compare results and interpretations.
2. Subgraph Patterns: Identify meaningful subgraphs (triangles, etc.) in a network. Analyze their distribution and link them to network properties.

Unit 4 **Teaching Hours: 12**

Graph-Based Machine Learning

Introduction to graph-based machine learning, Graph convolutional networks (GCNs) and their architecture.

Graph embedding techniques: node2vec, DeepWalk, Graph kernels and their applications, Graph neural networks for node classification and link prediction

Lab Exercise

1. Graph Embeddings: Use node2vec or DeepWalk to generate embeddings for a dataset. Visualize embeddings to explore relationships.
2. Graph Convolutional Networks with PyTorch Geometric: Build a simple GCN model for node classification on a citation network or similar dataset.

Unit 5 **Teaching Hours: 12**

Applications of Graph Analytics

Graph databases and querying graph data, Case studies in fraud detection and anomaly detection using graph analytics, Recommendation systems using graph-based algorithms, Graph analytics for biological networks and drug discovery, Ethical considerations and challenges in graph analytics.

Lab Exercise

1. Graph Database Exploration: Load a dataset into Neo4j. Learn basic Cypher queries and more advanced path-based queries.
2. Case Study Implementation: Work through significant parts of the fraud detection case study, including visualization, anomaly detection, potentially ML component.

Essential Reading

1. "Introduction to Graph Theory" by Richard J. Trudeau
2. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World" by David Easley and Jon Kleinberg
3. "Graph Algorithms" by Shimon Even and Guy Even
4. "Mining of Massive Datasets" by Jure Leskovec, Anand Rajaraman, and Jeffrey D. Ullman
5. "Networks: An Introduction" by Mark Newman

Recommended Reading

1. "Graph Theory and Complex Networks: An Introduction" by Maarten van Steen.
2. "Graph Analytics for Big Data: Applications, Algorithms, and Systems" by Charu C. Aggarwal.
3. "Mining Social Networks and Security Informatics" by Shishir Kumar Shandilya and Suresh Kumar Bodduluru.

MDS481: PROJECT-I (Web Project with Data Science Concepts)

Total Teaching Hours for Trimester: 60
No of hours per week: 5
Max Marks: 100 **Credits: 2**
Course Type: Major

Course Description
This course is designed to provide MSc Data Science students with hands-on experience in integrating data science techniques into web-based projects. In today's digital age, the web serves as a vast repository of data, presenting exciting opportunities for data scientists to extract insights, create impactful visualizations, and develop intelligent applications.

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

No.	Course Outcomes
CO1	Demonstrate proficiency in developing web applications
CO2	Show proficiency in Integrating Data Science Techniques with Web Development
CO3	Perform Effective Problem-solving and Decision-making Skills
CO4	Gain Advanced Understanding of Data Ethics and Best Practices

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS
UNIT 1	Identification of application domain to develop a web application/prototype demonstrating data science/ML methods.
UNIT 2	Requirement analysis, design and development of proposed solution using suitable tools for front end/data design.

Essential Reading

Recommended Reading
1. Flask Web Development: Developing Web Applications with Python, Second Edition, January 2018, Miguel Grinberg, O'reilly books
2. Django for Beginners: Build Websites with Python and Django, William S Vincent, welcometocode.
3. Interactive Data Visualization for the Web: An Introduction to Designing with D3, Second Edition, Scott Murray, O'reilly books.

MDS482: Research Problem Identification

Total Teaching Hours for Trimester: 30
No of hours per week: 3

Max Marks: 50 Credits: 1

Course Type: Major/Elective

Course Description

The objective of the course is to provide practical exposure to formal research paradigms in Data Science in various domains. Students apply research methodology principles to identify research based solution in their selected domains after a comprehensive literature review.

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

No.	Course Outcomes
CO1	Understand various data analysis paradigms used in various application domains
CO2	Identify gaps to propose research based solution

UNIT 1 Teaching Hours:30

Literature review in the identified research area, Study of Existing Model and Methodology, Research proposal development with a clearly defined Problem statement and a methodology for the implementation.

MDS571: BIG DATA ANALYTICS

Total Teaching Hours for Trimester: 75
No of hours per week: 7(4+3)
Max Marks: 100 **Credits: 4**
Course Type: Major

Course Description

The subject is intended to give the knowledge of Big Data evolving in every real-time application and how they are manipulated using the emerging technologies. This course breaks down the walls of complexity in processing Big Data by providing a practical approach to developing Java applications on top of the Hadoop platform. It describes the Hadoop architecture and how to work with the Hadoop Distributed File System (HDFS).

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

No.	Course Outcomes
CO1	Understand the Big Data concepts in real time scenario
CO2	Identify different types of Hadoop architecture
CO3	Demonstrate an ability to use Hadoop framework for processing Big Data for Analytics
CO4	Analyze the Big data under Spark architecture
CO5	Demonstrate the programming of Big data using Hive and Pig environments

Unit 1 **Teaching Hours: 15**

Introduction

Concepts of Data Analytics: Descriptive, Diagnostic, Predictive, Prescriptive analytics - Big Data characteristics: Volume, Velocity, Variety, Veracity of data - Types of data: Structured, Unstructured, Semi-Structured, Metadata - Introduction to Hadoop Scaling - Distributed Framework -Hadoop v/s RDBMS-Brief history of Hadoop.

Lab Exercise

1. Installing and Configuring Hadoop
2. Case study for identifying Data Characteristics

Unit 2 **Teaching Hours: 15**

Big Data Architecture

Standard Big data architecture - Big data application - Hadoop framework - HDFS Design goal - Master Slave architecture - Block System - Read-write Process for data - Installing HDFS - Executing in HDFS: Reading and writing Local files and Data streams into HDFS - Types of files in HDFS - Strengths and alternatives of HDFS - Concept of YARN. Apache Hadoop Moving Data in and out of Hadoop Understanding inputs and outputs of MapReduce - Problems with traditional large-scale systems-Requirements for a new approach.

Lab Exercise

1. Exercise on Reading and Writing Local files into HDFS
2. Exercise on Reading and Writing Data streams into HDFS

Unit 3 **Teaching Hours: 15**

Parallel Processing with MapReduce

Introduction to MapReduce - Sample MapReduce application: Wordcount - MapReduce Data types and Formats - Writing MapReduce Programming - Testing MapReduce Programs - MapReduce Job Execution - Shuffle and Sort - Managing Failures - Progress and Status Updates. MapReduce Programs: Using languages other than Java with Hadoop, Analyzing a large dataset.

Lab Exercise

1. Exercise on MapReduce applications
2. Exercise on writing and testing MapReduce Programs
3. Exercise on Shuffle and Sort
4. Exercise on Managing Failures

Unit 4 **Teaching Hours: 15**

Hive and Pig

Hive Architecture - Components - Data Definition - Partitioning - Data Manipulation - Joins, Views and Indexes - Hive Execution - Pig Architecture - Pig Latin Data Model - Latin Operators - Loading Data - Diagnostic Operators - Group Operators - Pig Joins - Row Level Operators - Pig Built-in function - User defined functions - Pig Scripts.

Lab Exercise

1. Exercise on Hive Architecture
2. Exercise on Pig Architecture

Unit 5 **Teaching Hours: 15**

Stream Processing with Spark

Stream processing Models and Tools - Apache Spark - Spark Architecture: Resilient Distributed Datasets, Directed Acyclic Graph - Spark Ecosystem - Spark for Big Data Processing: MLlib, Spark GraphX, SparkR, SparkSQL, Spark Streaming - Spark versus Hadoop . PySpark + NumPy + SciPy, Code Optimization.

Lab Exercise

1. Exercise on installing Spark
2. Exercise on Directed Acyclic Graph
3. Exercise on Spark using MLlib, Spark GraphX
4. Exercise on Spark using SparkR, Spark Streaming

Essential Reading

1. Anil Maheshwari (2020). *Big Data*. 2nd Edition. McGraw Hill Education Pvt Ltd.
2. S Chandramouli, Asha A George, C R Rene Robin,,D Doreen H Miriam,J JasmineC M, *Big Data Analytics*,University Press India Ltd., 2024

Recommended Reading

1. Thomas Erl, Wajid Khattak and Paul Buhler (2016). *Big Data Fundamentals: Concepts, Drivers and Techniques*. Service Tech Press.
2. Julián Luengo, Diego García-Gil, Sergio Ramírez-Gallego, Salvador García, Francisco Herrera (2020). *Big Data Preprocessing: Enabling Smart Data*. Springer Nature Publishing.
3. Seema Acharya, Subhasini Chellappan (2019), *Big Data and Analytics*. 2nd Edition, Wiley India Pvt Ltd

MDS531A: ECONOMETRICS

Total Teaching Hours for Trimester: 60
No of hours per week: 5
Max Marks: 100 **Credits: 4**
Course Type: Elective

Course Description

The course is designed to impart the learning of principles of econometric methods and tools. This is expected to improve student's ability to understand of econometrics in the study of economics and finance. The learning objective of the course is to provide students to get the basic knowledge and skills of econometric analysis, so that they should be able to apply it to the investigation of economic relationships and processes, and also understand the econometric methods, approaches, ideas, results and conclusions met in the majority of economic books and articles. Introduce the students to the traditional econometric methods developed mostly for the work with cross-sections data.

Course Outcome

No.	Course Outcomes
CO1	Demonstrate Simple and multiple Econometric models
CO2	Interpret the models adequacy through various methods
CO3	Demonstrate simultaneous Linear Equations model.
CO4	Demonstrate contemporary trends in estimation of econometrics models.

Unit 1 **Teaching Hours: 15**

Introduction to Econometrics

Introduction to Econometrics- Meaning and Scope – Methodology of Econometrics – Nature and Sources of Data for Econometric analysis – Types of Econometrics, scope and limitations of econometrics, Generalised Least Squares (GLS) Estimator.

Unit 2 **Teaching Hours: 15**

Econometric Models and Their Inference

Presence of outliers, omitted variables, nonlinear relationship, correlated disturbances heteroscedasticity. Linear Regression with Stochastic Regressors, Errors in Variable Models and Instrumental Variable Estimation, Independent Stochastic linear Regression, Auto regression, Linear regression, Lag Models.

Unit 3 **Teaching Hours: 15**

Linear Equations Model

Simultaneous Linear Equations Model: Structure of Linear Equations Model, Identification Problem, Rank and Order Conditions, Single Equation and Simultaneous Equations, Methods of Estimation- Indirect Least squares, Least Variance Ratio and Two Stage Least Square, reduced form method or indirect least squares (ILS), the method of instrumental variables (IV), two-stage least squares (2SLS).

Unit 4 **Teaching Hours: 15**

Dummy Variable Regression Models

Meaning and Nature of dummy variables, ANOVA models, dummy variable alternative to the chow test, interaction effects using dummy variables, the use of dummy variables in seasonal analysis, piecewise linear regression, panel data regression models, dummy variables and heteroscedasticity, dummy variables and autocorrelation.

Essential Reading

1. Gujarati, D. N. (2021). Essentials of econometrics. Sage Publications.
2. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons.
3. Dinardo, J., Johnston, J., & Johnston, J. (1997). Econometric methods. McGraw-Hill Companies, Inc.

Recommended Reading

1. Intriligator, M. D. (1980). Econometric Models-Techniques and Applications, Prentice Hall.
2. Theil, H. (1971). Principles of Econometrics, John Wiley.
3. Walters, A. (1970). An Introduction to Econometrics, McMillan and Co.

MDS531B: Bayesian Inference

Total Teaching Hours for Trimester: 60
No of hours per week: 5
Max Marks: 100 **Credits: 4**
Course Type: Elective
Course Description and Course Objectives

Students who complete this course will gain a solid foundation in how to apply and understand Bayesian statistics and how to understand Bayesian methods vs frequentist methods. Topics covered include: an introduction to Bayesian concepts; Bayesian inference for binomial proportions, and normal means; modelling.

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

No.	Course Outcomes
CO1	Identify Bayesian methods for a binomial proportion.
CO2	Analyse normal distributed data in the Bayesian framework.
CO3	Compare Bayesian methods and frequentist methods.

Unit 1 **Teaching Hours: 12**
Introduction to Bayesian Thinking

Basics of minimaxity - subjective and frequentist probability - Bayesian inference - prior distributions - posterior distributions - loss function - the principle of minimum expected posterior loss - quadratic and other common loss functions - advantages of being Bayesian - Improper priors - common problems of Bayesian inference - Point estimators - Bayesian confidence intervals, testing – credible intervals.

Unit 2 **Teaching Hours: 12**
Bayesian Inference for Discrete Random Variables

Two Equivalent Ways of Using Bayes' Theorem - Bayes' Theorem for Binomial with Discrete Prior-Important Consequences of Bayes' Theorem - and Bayes' Theorem for Poisson with Discrete prior.

Unit 3 **Teaching Hours: 12**
Bayesian Inference for Binomial Proportion

Using a Uniform Prior - Using a Beta Prior - Choosing Your Prior - Summarizing the Posterior Distribution - Estimating the Proportion - Bayesian Credible Interval - Statistical inference from both frequentist and Bayesian perspectives-Hypothesis Testing - Testing a One-Sided Hypothesis - Testing a Two-Sided Hypothesis.

Unit 4 **Teaching Hours: 12**
Bayesian Inference for Normal Mean

Bayes' Theorem for Normal Mean with a Discrete Prior - Bayes' Theorem for Normal Mean with a Continuous Prior - Normal Prior, Bayesian Credible Interval for Normal Mean - Predictive Density for Next Observation.

Unit 5 **Teaching Hours: 12**
Bayesian Computations

Analytic approximation - E-M Algorithm - Monte Carlo sampling - Markov Chain Monte Carlo Methods - Metropolis-Hastings Algorithm - Gibbs sampling: examples and convergence issues - Bayesian linear regression.

Essential Reading

1. Bolstad W. M. and Curran, J.M. (2016) Introduction to Bayesian Statistics 3rd Edition. Wiley, New York
2. Jim, A. (2009). Bayesian Computation with R, 2nd Edition, Springer.

Recommended Reading

1. Berger, J.O. (1985a). Statistical Decision Theory and Bayesian Analysis, 2nd Ed. Springer-Verlag, New York.
2. Christensen R, Johnson, W., Branscum, A. and Hanson T. E. (2011). Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians, Chapman & Hall.
3. Congdon, P. (2006). Bayesian Statistical Modeling, Wiley
4. Ghosh, J. K., Delampady M. and T. Samantha (2006). An Introduction to Bayesian Analysis: Theory & Methods, Springer.
5. Rao. C.R. and Day. D. (2006). Bayesian Thinking, Modeling & Computation, Handbook of Statistics, Vol. 25. Elsevier.

MDS531C: BIOSTATISTICS

Total Teaching Hours for Trimester: 60
No of hours per week: 5
Max Marks: 100 **Credits: 4**
Course Type: Elective

Course Description
This course provides an understanding of various statistical methods in describing and analyzing biological data. Students will be equipped with an idea about the applications of statistical hypothesis testing, related concepts and interpretation in biological data.
Course Outcomes: Upon Completion of the course students will be able to

No.	Course Outcomes
CO1	Demonstrate the understanding of basic concepts of biostatistics and the process involved in the scientific method of research.
CO2	Identify how the data can be appropriately organized and displayed.
CO3	Analyze and interpret the data based on the discrete and continuous probability distributions. Apply parametric and non-parametric methods of statistical data analysis.
CO4	Understand the concepts of Epidemiology and Demography

Unit 1 **Teaching Hours: 12**
Introduction to Biostatistics
Types of variables. Measurement and measurement levels of the variables in biological science. Visualization and descriptive analysis of biological data. Sensitivity, Specificity, Positive predictive value, Negative predictive value. ROC Curves.

Unit 2 **Teaching Hours: 12**
Parametric and Non - Parametric Methods
Parametric methods in biological data analysis: One sample t-test - independent sample t-test - paired sample t-test - one-way analysis of variance - two-way analysis of variance - analysis of covariance - repeated measures of analysis of variance, Post Hoc Analysis for ANOVA, Pearson correlation coefficient: Introduction to non- parametric methods and its use in biological data.

Unit 3 **Teaching Hours: 12**
Generalized Linear Models
Review of simple and multiple linear regression - introduction to generalized linear models - parameter estimation of generalized linear models - models with different link functions - binary (logistic) regression - estimation and model fitting - Poisson regression for count data - mixed effect models and hierarchical models with practical examples.

Unit 4 **Teaching Hours: 12**
Basics of Epidemiology
Introduction to epidemiology, measures of epidemiology, observational study designs: case report, case series correlational studies, cross-sectional studies, retrospective and prospective studies, analytical epidemiological studies-case control study and cohort study, odds ratio, relative risk, the bias in epidemiological studies.

Unit 5 **Teaching Hours: 12**

Demography

Introduction to demography, mortality and life tables, infant mortality rate, standardized death rates, life tables, fertility, crude and specific rates, migration-definition and concepts population growth, measurement of population growth-arithmetic, geometric and exponential, population projection and estimation, different methods of population projection, logistic curve, urban population growth, components of urban population growth.

Essential Reading

1. Rosner, B. A. (2011). Fundamentals of Biostatistics. Austria: Brooks/Cole
2. Leon Gordis, Epidemiology

Recommended Reading

1. Marcello Pagano and Kimberlee Gauvreau (2018), Principles of Biostatistics, 2nd Edition, Chapman and Hall/CRC press.
2. Park K., (2019), *Park's Text Book of Preventive and Social Medicine,* Banarsidas Bhanot, Jabalpur

MDS572A-EVOLUTIONARY ALGORITHMS

Total Teaching Hours for Semester:60
No of Lecture Hours/Week: 6 (3+3) **Credits:3**
Max Marks:100

Course Description and Course Objectives
Able to understand the core concepts of evolutionary computing techniques and popular evolutionary algorithms that are used in solving optimization problems. Students will be able to implement custom solutions for real-time problems applicable with evolutionary computing.

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

No.	Course Outcomes
CO	Basic understanding of evolutionary computing concepts and techniques.
CO2	Classify relevant real-time problems for the applications of evolutionary algorithms.
CO3	Design solutions using evolutionary algorithms.

Unit 1 **Teaching Hours: 12**
Introduction to Evolutionary Computing
Terminologies – Notations – Problems to be solved – Optimization – Modeling – Simulation – Search problems – Optimization constraints
Lab Exercise
1. Implementation of single and multi-objective functions
2. Implementation of binaryGA

Unit 2 **Teaching Hours: 12**
Evolutionary Programming
Continuous evolutionary programming – Finite state machine optimization – Discrete evolutionary programming – The Prisoner's dilemma
STRATEGY: One plus one evolution strategy – The 1/5 Rule – (μ+1) evolution strategy – Self adaptive evolution strategy
Lab Exercise
1. Implementation of continuousGA.
2. Implementation of evolutionary programming.

Unit 3 **Teaching Hours: 12**
Genetic Programming
BASICS: Fundamentals of genetic programming – Genetic programming for minimal time control. EVOLUTIONARY ALGORITHM VARIATION: Initialization – Convergence – Population diversity – Selection option – Recombination – Mutation.
Lab Exercise
1. Implementation of geneticprogramming
2. Implementation of Ant ColonyOptimization

Unit 4 **Teaching Hours: 12**
Optimization Models
ANT COLONY OPTIMIZATION: Pheromone models – Ant system – Continuous Optimization – Other Ant System. PARTICLE SWARM OPTIMIZATION: Velocity limiting – Inertia weighting – Global Velocity updates – Fully informed Particle Swarm.

Lab Exercise

1. Implementation of Particle SwarmOptimization
2. Implementation of Multi-ObjectOptimization

Unit 5 **Teaching Hours: 12**

Mult-Objective Optimiation

Pareto Optimality – Hyper volume – Relative coverage – Non-pareto based EAs – Pareto based EAs – Multi-objective Biogeography based optimization

Lab Exercise

1. Simulation of EA in Planning problems (routing, scheduling, packing) and Design problems (Circuit, structure,art)
2. Simulation of EA in classification/predictionmodelling

Essential Reading

1. D. Simon, Evolutionary optimization algorithms: biologically inspired and population-based approaches to computer intelligence. New Jersey: John Wiley, 2013.
2. Eiben and J. Smith, Introduction to evolutionary computing. 2nd ed. Berlin: Springer, 2015.

Recommended Reading

1. D.Goldberg, Genetic algorithms in search, optimization, and machine learning. Boston: Addison-Wesley, 2012.
2. K. Deb, Multi-objective optimization using evolutionary algorithms. Chichester: John Wiley & Sons, 2009.

MDS572B: QUANTUM MACHINE LEARNING

Total Teaching Hours for Trimester: 60 Hrs.
No of hours per week: 6(3+3) **Credits: 3**
Max Marks: 100
Course Type: Elective

Course Description

This course explores the intersection of quantum computing and machine learning, introducing students to the fundamental principles of quantum mechanics and their application in designing quantum algorithms for machine learning tasks. Students will gain hands-on experience in implementing quantum machine learning algorithms using relevant programming frameworks. The course aims to equip students with the knowledge and skills necessary to navigate the rapidly evolving field of quantum machine learning.

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

No.	Course Outcomes
CO1	Understand the basics of quantum mechanics and quantum computing.
CO2	Implement and analyze quantum machine learning algorithms using Qiskit
CO3	Apply quantum algorithms to solve machine learning problems.
CO4	Critically evaluate the advantages and limitations of quantum machine learning approaches.

Unit 1 **Teaching Hours: 12**
Introduction to Quantum Mechanics (2 Weeks)

Introduction and overview, Global perspectives, Quantum bits, Quantum computation, Quantum algorithms, Quantum information processing.
Introduction to Quantum Mechanics - The postulates of Quantum Mechanics, Application: superdense coding, The density operator.

LAB Exercise

1. Install Qiskit and set up the development environment.

Unit 2 **Teaching Hours: 12**
Introduction to Quantum Computation (2 Weeks)

Quantum Circuits - Quantum algorithms, Single Qubit operations, Controlled operations, Measurement, Universal Quantum gates. Simulation of Quantum systems.

LAB Exercise

2. Basic operations on Qubit and measurements on Bloch Sphere
3. Create a simple quantum circuit using basic gates
4. Visualize and simulate the quantum circuit using Qiskit
5. Quantum Solution to the Deutsch-Josza Problem

Unit 3 **Teaching Hours: 12**
Clustering Structure and Quantum Computing (2 Weeks)

Quantum Random Access Memory, Quantum Principal Component Analysis, Quantum K-Means, Quantum Hierarchical Clustering.

LAB Exercise

6. Implement a quantum clustering algorithm using Qiskit or a similar library.
7. Apply the quantum algorithm to a dataset and visualize the cluster structure.

Unit 4 **Teaching Hours: 12**

Quantum Classification (2 Weeks)

Nearest Neighbors, Support Vector Machines with Grover's Search, Support Vector Machines with Exponential Speedup, Computational Complexity

LAB Exercise

8. Implement Quantum Kernels and Support Vector Machines
9. Design a Training Parameterized Quantum Circuits

Unit 5 **Teaching Hours: 12**

Quantum Pattern Recognition (2 Weeks)

Quantum Associative Memory, The Quantum Perceptron, Quantum Neural Networks, Physical Realizations. Variational quantum algorithms and their applications

LAB Exercise

10. Implement a simple quantum neural network and evaluate the performance

Essential Reading

1. Quantum Computation and Quantum Information by Michael Nielsen and Isaac Chuang
2. Quantum Machine Learning: What Quantum Computing Means to Data Mining by Peter Wittek

Recommended Reading

1. Quantum Computing for Computer Scientists by Noson S. Yanofsky and Mirco A. Mannucci
2. Quantum Machine Learning: A Gentle Introduction by Jacob Biamonte, Peter Wittek, and Nicola Pancotti
3. Quantum Machine Learning: Theory and Experiments" by Maria Schuld and Francesco Petruccione
4. Learn Quantum Computing with Python and Q# by Sarah C. Kaiser and Christopher Granade"

Online learning and Lab References:

- Introduction to Quantum Computing: Quantum Algorithms and Qiskit - Course (nptel.ac.in)
- **IBM Quantum Learning**
- **Qiskit Foundations - Coding with Qiskit - YouTube**
- Quantum Machine Learning - Qiskit - YouTube

MDS572C: REINFORCEMENT LEARNING

Total Teaching Hours for Trimester: 60 Hrs.
No of hours per week: 6(3+3)
Max Marks: 100 **Credits: 3**
Course Type: Elective
Course Description

The main objective of this course is to teach students how to define reinforcement learning problems and apply algorithms such as dynamic programming, Monte Carlo, and temporal-difference learning to solve them. Students will advance towards more complex state space environments by employing function approximation, deep Q-networks, and cutting-edge policy gradient techniques. We will also discuss current approaches rooted in reinforcement learning, including imitation learning, meta learning, and more intricate environment formulations.

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

No.	Course Outcomes
CO1	Grasp the fundamental concepts of Reinforcement Learning, including Markov Decision Processes, states, actions, rewards, and key components of RL System.
CO2	Able to apply dynamic programming methods
CO3	Develop skills in model-free prediction using Monte Carlo methods.
CO4	Comprehend various exploration strategies such as epsilon-greedy, softmax exploration, and Upper Confidence Bound (UCB)
CO5	Apply and understand Policy Gradient method in Reinforcement Learning Environment

Unit 1 **Teaching Hours: 12**
Introduction to Reinforcement Learning

Overview of Reinforcement Learning (RL)-Definition and key concepts - Contrasting RL with supervised and unsupervised learning, Markov Decision Processes (MDPs) - States, actions, and rewards - Transition probabilities and dynamics - Bellman equation and optimality. Value Functions - State-value and action-value functions.

Lab Exercise

1. Hands-on activity: participants formalize a simple problem as an MDP.
2. Familiarize students with the OpenAI Gym library for reinforcement learning.
3. Set up OpenAI Gym and create a simple environment

Unit 2 **Teaching Hours: 12**
Dynamic Programming Approaches

Policy Evaluation, Policy Improvement, Policy iteration, Value iteration, Asynchronous Dynamic Programming

Lab Exercises

1. Create a Tic-Tac-Toe Game Using RL

Unit 3 **Teaching Hours: 12**
Model-Free Prediction

Monte Carlo methods, Temporal Difference (TD) learning, Eligibility traces.

Lab Exercise

1. Design and implement a Monte Carlo algorithm for solving a specific environment, and discuss the key components of the algorithm, such as episode generation, state-value estimation, and policy improvement.

2. Implement the Temporal Difference (TD) prediction algorithm (e.g. Q-learning

Unit 4 **Teaching Hours: 12**

Exploration and Exploitation

Exploration Strategies-Epsilon-greedy- Softmax exploration - Upper Confidence Bound (UCB), Multi-Armed Bandits

Lab Exercise

1. Implementing an Adversarial Bandit algorithm in python framework.

Unit 5 **Teaching Hours: 12**

Policy Gradient Method

Policy Approximation and its advantages,Actor-Critic Methods - Advantage functions- A3C (Asynchronous Advantage Actor-Critic),Policy Gradient for continuous problem.

LAB Exercises

1. Building a simple Actor-Critic model for a basic environment (e.g., CartPole)
2. Implement a state-of-the-art policy optimization algorithm.
3. Implement an Actor-Critic algorithm for continuous action space

Essential Reading

1. Richard S. Sutton and Andrew G. Barto, "Reinforcement learning: An introduction", Second Edition, MIT Press, 2019
2. Dimitri Bertsekas and John G. Tsitsiklis, Neuro Dynamic Programming, Athena Scientific. 1996. ISBN-13: 978-1886529106

Recommended Reading

1. V. S. Borkar, Stochastic Approximation: A Dynamical Systems Viewpoint, Hindustan Book Agency, 2009. ISBN-13: 978-0521515924
2. Deep Learning. Ian Goodfellow and Yoshua Bengio and Aaron Courville. MIT Press. 2016.ISBN-13: 978-0262035613.

MDS573A: GEO-SPATIAL DATA ANALYTICS

Total Teaching Hours for Trimester: 60
No of hours per week:6(3+3)
Max Marks: 100 **Credits: 3**
Course Type: Elective

Course Description
This course aims to provide students with a comprehensive understanding of geospatial data analytics techniques, tools and applications. Students will learn to analyze, interpret, and visualize spatial data to derive meaningful insights.

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

No.	Course Outcomes
CO1	Understand fundamental geospatial data analysis techniques
CO2	Apply geospatial data visualization methods to represent spatial patterns and trends
CO3	Apply different geospatial analysis techniques .
CO4	Implement geospatial data analytics workflows using relevant software tools.

Unit 1 **Teaching Hours: 12**
Introduction to Geospatial Data Analytics
Overview of geospatial data and its sources, Introduction to GIS (Geographic Information Systems), spatial data structures, coordinate systems and data format. Basic spatial data analysis techniques, spatial querying, buffering and overlay operations.
Lab Exercise Introduction to ArcGIS or QGIS to load spatial data layers and perform basic spatial analysis tasks.

Unit 2 **Teaching Hours: 12**
Spatial Data Visualization
Principles and techniques of spatial data visualization , Cartography and map design.Visualization techniques for spatial data : choropleth maps, proportional symbol maps and heatmaps. interactive mapping tools and platforms for creation of dynamic, web-based maps.
Lab Exercise: Creating thematic maps using ArcGIS Online or Carto.
Unit 3 **Teaching Hours: 12**
Spatial Analysis Techniques Teaching
Introduction to Spatial interpolation methods: inverse distance weighting and kriging. Spatial clustering and pattern analysis. Understanding Geostatistics and spatial regression

Lab Exercise: Spatial analysis for automation of geospatial workflows using Python with libraries like GeoPandas and PySAL

Unit 4 **Teaching Hours: 12**
Geospatial Data Mining and Machine Learning
Introduction to geospatial data mining. Understanding Machine learning algorithms like decision trees, random forests, and support vector machines for spatial data analysis Land cover classification, spatial prediction and anomaly detection.
Lab Exercise: Implementing machine learning models for spatial prediction tasks.

Unit 5 **Teaching Hours: 12**

Advanced Topics in Geospatial Data Analytics

Understanding Big data analytics for geospatial data, Web mapping and spatial data APIs Introduction to Spatial data integration and interoperability

Lab Exercise: Building a web-based geospatial application using Leaflet.js

Essential Reading

1. Joel Lawhead, Learning Geospatial Analysis with Python, Fourth Edition, Packt Publishing, 2023.
2. Michael J. de Smith, Michael F. Goodchild, and Paul A. Longley, Geospatial Analysis: A Comprehensive Guide, Winchelsea Press,2018

Recommended Reading

1. Paul Bolstad ,GIS Fundamentals: A First Text on Geographic Information Systems, XanEdu Publishing Inc, 2020
2. Aurelia Moser, Jon Bruner, Bill Day, Geospatial Data and Analysis, O'Reilly Media Inc, 2017

Web Links

1. Esri Training: <https://www.esri.com/en-us/training/>
2. GIS Lounge: <https://www.gislounge.com/>

MDS573B: BIOINFORMATICS

Total Teaching Hours for Trimester: 60
No of hours per week: 6(3+3)
Max Marks: 100 **Credits: 3**
Course Type: Elective

Course Objectives

1. Provide an overview of the Machine Learning concepts and practices in Bioinformatics
2. Gain experience in applications and limitations of Machine Learning
3. To encompass a broad range of approaches to data analysis across the biological sciences

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

No.	Course Outcomes
CO1	Understand how to evaluate models generated from data
CO2	Understand public-domain biological datasets
CO3	Analyze genomics using decision trees, and random forests
CO4	Design computational experiments for training and evaluating machine learning methods for solving bioinformatics problems

Unit 1 **Teaching Hours: 12**
Introduction to Bio-Informatics Data and Databases
Types of Biological data:-genomic DNA, Complementary DNA, Recombinant DNA, Expressed sequence tags, Sequence -Tagged sites.
Lab Exercise

1. Create directories and verify the directory commands.
2. Create the file(s) and verify the file handling commands.
3. Retrieval of Data from Biological Database.

Unit 2 **Teaching Hours: 12**
Gene Selection using Omics Data
Approaches for Gene selection - multi-level omics data integration, Machine learning approaches for multi-level data integration, Random Forest algorithm in imbalance genomics classification
Lab Exercise:

1. Protein Sequence Retrieval from Uniprot.
2. Global and Local Alignment.
3. Dot Plot Sequence alignment.

Unit 3 **Teaching Hours: 12**
Microarray Data Optimization
Microarray data, Grey Wolf Optimization (GWO) Algorithm, Studies on GWO variants, Application of GWO in medical domain, Application of GWO in Microarray data. Case study**,** Using AI to detect Coronavirus. **Healthcare Solutions**: **U**sing machine learning approaches for different purposes, Various resources of medical data set for research, Deep learning in Health care, Projects in medical imaging and diagnostics**.**
Lab Exercise:

1. Retrieve genetic sequence data using BLAST(Basic Local Alignment Search Tool
2. Protein secondary structure prediction
3. Protein 3D structure visualization

Unit 4 **Teaching Hours: 12**

Python for Bioinformatics working with BioPython

Representing sequence data: Storing DNA sequence, Concatenating DNA fragments, Transcription DNA to RNA, Proteins,Files and Arrays, Reading Proteins in Files, Arrays, dictionary and List Context.

Lab Exercise

1. Read protein sequence data from a file.
2. Search for a motif in a DNA sequence.

Unit 5 **Teaching Hours: 12**

The Genetic Code

GenBank: GenBank files, GenBank Libraries, Separating Sequence and Annotation, AParsing Annotations, Indexing GenBank with DBM. Protein Data Bank: Files and Folders, PDB Files, Parsing PDB Files.

Lab Exercise

1. Case Study:

- To retrieve the sequence of the Human keratin protein from UniProt database and to interpret the results.
- To retrieve the sequence of the Human keratin protein from the GenBank database and to interpret the results.

Essential Reading

1. S.C. Rastogi et al. Bioinformatics: Methods and Applications: (Genomics, Proteomics and Drug Discovery) Kindle Edition.(UNIT I)
2. Data Analytics in Bioinformatics: A Machine Learning Perspective by Rabinarayan Satpathy, Xiaobo Zhang, Sachi Nandan Mohanty, Suneeta Satpathy, Tanupriya Choudhury, 2021, John Wiley & Sons. (UNIT2,UNIT3)
3. Taneja & Kumar: Python Programming: A Modular Approach, Pearson Kenneth & Lambert: Fundamental of Python. Course Technology Chang, Chapman, et al. Biopython Tutorial and Cookbook (ebook).(UNIT4)

Recommended Reading

1. Stuart Russel and Peter Norvig, "Artificial Intelligence- A Modern Approach", Prentice Hall, 1995.
2. Bioinformatics Technologies, Yi-Ping Phoebe Chen (Ed), 1st edition, Springer, 2005.
3. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems, by Aurelien Geron, 2019, O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
4. Introduction to Bioinformatics, by Arthur Lesk, 5th Edition, 2019, Oxford University Press, UK.

Web resources:

[1] https://canvas.harvard.edu/courses/8084/assignments/syllabus
[2] https://www.coursera.org/specializations/bioinformatics
[3] http://www.dtc.ox.ac.uk/modules/introduction-bioinformatics-bioscientists.html

MDS573C: IMAGE AND VIDEO ANALYTICS

Total Teaching Hours for Trimester: 60
No of hours per week: 6(3+3)
Max Marks: 100 **Credits: 3**
Course Type: Elective
Course Description
This course will provide a basic foundation towards digital image processing and video analysis. This course will also provide a brief introduction about various Object Detection, Recognition, Segmentation and Compression methods which will help the students to demonstrate real-time image and video analytics applications.
Course Outcomes: Upon completion of the course students will be able to

No.	Course Outcomes
CO1	Understand the fundamental principles of image and video analysis
CO2	Develop proficiency in image enhancement and segmentation
CO3	Develop skills in object detection and recognition
CO4	Apply the image and video analysis approaches to solve real world problems

Unit 1 **Teaching Hours: 12**
Introduction to Digital Image and Video Processing
Digital image representation, Sampling and Quantization, Types of Images, Basic Relations between Pixels - Neighbors, Connectivity, Distance Measures between pixels, Introduction to Digital Video, Sampled Video, Video Transmission. Gray-Level Processing: Image Histogram, Linear and Non-linear point operations on Images, Image Thresholding, Region labelling, Binary Image Morphology.
Lab Exercise
1. Adjust the threshold of the image and analyze the area of a specific component in the image.
2. Program to implement contrast stretching.

Unit 2 **Teaching Hours: 12**
Image and Video Enhancement and Restoration
Spatial domain-Linear and Non-linear Filtering, Introduction to Fourier Transform and the frequency Domain– Filtering in Frequency domain, A model of The Image Degradation /Restoration, Noise Models and basic methods for image restoration.
Lab Exercise
1. Program to analyze and image to find total number, total area, average size of elements
2. Program to implement Non-linear Spatial Filtering using Built-in and user defined functions.

Unit 3 **Teaching Hours: 12**
Image and Video Compression
Fundamentals of Image Compression: Huffman Coding, Run length Coding, LZW Coding, Bit plane coding. Video Compression: Basic Concepts and Techniques of Video compression, MPEG-1 and MPEG-2 Video Standards.
Lab Exercise
1. Program to compare performance of various image compression methods.
2. Program to Extract frames from videos and analyze each frame.

Unit 4 **Teaching Hours: 12**

Feature Detection and Description

Introduction to feature detectors, Point, line and edge detection, Image Segmentation - Region Based Segmentation – Region Growing and Region Splitting and Merging, Thresholding – Basic global thresholding, optimum global thresholding using Otsu's Method.

Lab Exercise

1. Find out the number of labeled components in an image. Also find the area and integrated density of the component.
2. Analyze the morphology of specific components in the given image.

Unit 5 **Teaching Hours: 12**

Object Detection and Recognition

Descriptors: Boundary descriptors - Fourier descriptors - Regional descriptors

Object detection and recognition in image and video: Minimum distance classifier, Applications in image and video analysis, object tracking in videos.

Lab Exercise

1. Extracting feature descriptors from the image dataset.
2. Implement object tracking in videos.

Essential Reading

1. Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, 4th Edition, Pearson Education, 2018.
2. Alan Bovik, Handbook of Image and Video Processing, Second Edition, Academic Press, 2005.

Recommended Reading

1. Anil K Jain, Fundamentals of Digital Image Processing, PHI, 2011.
2. Richard Szeliski, Computer Vision Algorithms and Applications, Springer,2011.
3. Oge Marques, Practical Image and Video Processing Using MatLab, Wiley, 2011.
4. John W. Woods, Multidimensional Signal, Image, Video Processing and Coding, Academic Press, 2006.

Unit 2 **Teaching Hours: 12**

Gene Selection using Omics Data

Approaches for Gene selection - multi-level omics data integration, Machine learning approaches for multi-level data integration, Random Forest algorithm in imbalance genomics classification

Lab Exercise:

4. Protein Sequence Retrieval from Uniprot.
5. Global and Local Alignment.
6. Dot Plot Sequence alignment.

Unit 3 **Teaching Hours: 12**

Microarray Data Optimization

Microarray data, Grey Wolf Optimization (GWO) Algorithm, Studies on GWO variants, Application of GWO in medical domain, Application of GWO in Microarray data. Case study, Using AI to detect Coronavirus. **Healthcare Solutions**: Using machine learning approaches for different purposes, Various resources of medical data set for research, Deep learning in Health care, Projects in medical imaging and diagnostics.

Lab Exercise:

4. Retrieve genetic sequence data using BLAST(Basic Local Alignment Search Tool
5. Protein secondary structure prediction

6. Protein 3D structure visualization

Unit 4 **Teaching Hours: 12**

Python for Bioinformatics working with BioPython

Representing sequence data: Storing DNA sequence, Concatenating DNA fragments, Transcription DNA to RNA, Proteins,Files and Arrays, Reading Proteins in Files, Arrays, dictionary and List Context.

Lab Exercise

3. Read protein sequence data from a file.
4. Search for a motif in a DNA sequence.

Unit 5 **Teaching Hours: 12**

The Genetic Code

GenBank: GenBank files, GenBank Libraries, Separating Sequence and Annotation, AParsing Annotations, Indexing GenBank with DBM. Protein Data Bank: Files and Folders, PDB Files, Parsing PDB Files.

Lab Exercise

1. Case Study:
   - To retrieve the sequence of the Human keratin protein from UniProt database and to interpret the results.
   - To retrieve the sequence of the Human keratin protein from the GenBank database and to interpret the results.

Essential Reading

4. S.C. Rastogi et al. Bioinformatics: Methods and Applications: (Genomics, Proteomics and Drug Discovery) Kindle Edition.(UNIT I)
5. Data Analytics in Bioinformatics: A Machine Learning Perspective by Rabinarayan Satpathy, Xiaobo Zhang, Sachi Nandan Mohanty, Suneeta Satpathy, Tanupriya Choudhury, 2021, John Wiley & Sons. (UNIT2,UNIT3)
6. Taneja & Kumar: Python Programming: A Modular Approach, Pearson Kenneth & Lambert: Fundamental of Python. Course Technology Chang, Chapman, et al. Biopython Tutorial and Cookbook (ebook).(UNIT4)

Recommended Reading

5. Stuart Russel and Peter Norvig, “Artificial Intelligence- A Modern Approach”, Prentice Hall, 1995.
6. Bioinformatics Technologies, Yi-Ping Phoebe Chen (Ed), 1st edition, Springer, 2005.
7. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems, by Aurelien Geron, 2019, O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
8. Introduction to Bioinformatics, by Arthur Lesk, 5th Edition, 2019, Oxford University Press, UK.

Web resources:

[1] https://canvas.harvard.edu/courses/8084/assignments/syllabus
[2] https://www.coursera.org/specializations/bioinformatics
[3] http://www.dtc.ox.ac.uk/modules/introduction-bioinformatics-bioscientists.html

MDS581: PROJECT II (RESEARCH PROJECT/CAPSTONE PROJECT)

Total Teaching Hours for Trimester: 60
No of hours per week: 5
Max Marks: 100 Credits: 2
Course Type: Major

Course Description

The Capstone/Research Project in Data Science provide students with the opportunity to integrate and apply the knowledge and skills acquired throughout their coursework to address real-world data science challenges.
This course emphasizes advanced research methodologies, data analysis techniques, and effective communication of findings.
Students will work individually or in teams of 2 under the supervision of faculty advisors to complete a substantial research project. Projects may focus on a wide range of topics within the field of data science, including but not limited to machine learning, data mining, natural language processing, computer vision, predictive modelling big data analytics etc.

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

No.	Course Outcome`
CO1	To demonstrate advanced proficiency in conducting independent research in the field of data science,
CO2	To apply advanced statistical and machine learning techniques to analyze complex datasets
CO3	Students will develop and apply creative problem-solving skills to address data science challenges,
CO4	Students will develop proficiency in project management skills

Unit:1
Identifying a research question or problem statement relevant to the field of data science, Conducting a comprehensive literature review to understand the existing research and methodologies related to the chosen topic, designing and implementing appropriate data collection and pre-processing techniques,

Unit:2
Applying advanced statistical/Data Science/machine learning algorithms to analyse and interpret the data, Evaluating the performance of the models and methodologies employed, Drawing meaningful conclusions and insights from the analysis, Presenting the research findings through written reports, oral presentations, and potentially through visualization or demonstration of prototypes

MDS682: Research Publication

Total Teaching Hours for Trimester: 30
No of hours per week: 3
Max Marks: 50 Credits: 2
Course Type: Major

Course Description
The objective of the course is to provide practical exposure to major data analysis paradigms in various

application domains for performing research. Students complete the implementation of identified research problem and present the finding through a research paper.

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

No.	Course Outcomes
CO1	Analyze various data science paradigms
CO2	Build a data science model to provide solution to the identified problem

UNIT 1 Teaching Hours:30

Implementation of proposed solution with relevant tools. Writing a research paper to present the findings and communication to identified journal.