CHRIST
(DEEMED TO BE UNIVERSITY)
PUNE LAVASA CAMPUS
The Hub of Analytics

SYLLABUS

MSC

DATA SCIENCE (2023-24)

CHRIST (Deemed to be University)
Pune Lavasa Campus - 'The Hub of Analytics'

1800 123 2009
lavasa.christuniversity.in

Department of Statistics and Data Science

Syllabus

MSc (Data Science)

AY 2023-24

CHRIST (Deemed to be University), Bangalore.

Karnataka, India

Syllabus for Master of Science (Data Science) 2023-24 approved by the Board of Studies, Department of Statistics and Data Science and Academic Council, CHRIST (Deemed to be University), Bangalore, India.

Published by the Centre for Publications, CHRIST (Deemed to be University), Hosur Road, Bangalore, 560 029, India. publications@christuniversity.in

2023

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 four semesters. 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 two years of learning)

1. Computer Science
2. Mathematics
3. Statistics

PROGRAMME STRUCTURE:

TRIMESTER-I

Course Code	Course Title	Course hrs	Hours Per Week	Credits	Marks
MDS131	Research methods in Data Science	60	5	4	100
MDS132	Probability and Distribution Theory	60	5	4	100
MDS133	Mathematical Foundations for Data Science-I	45	4	3	100
	Choose Any One (Foundational Elective)				
MDS161A	Foundation Elective-I (Principles of Programing)	30	3	2	50
MDS161B	Foundation Elective-II (Introduction to Probability and Statistics)	30			
MDS161C	Foundation Elective-III(Linux Essentials)	30			
MDS171	Programming using Python	90	8 (4+4)	5	150
MDS151	Applied Excel	30	3	1	50
HOLODD	HOLISTIC EDUCATION		1	1	50
Total	**-**		**29**	**20**	**550**

TRIMESTER-II

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

TRIMESTER-III

Course Code	Course Title	Course hours	Hours Per Week	Credits	Marks
MDS331	Regression Modelling	45	4	3	100
MDS371	Java Programming	75	7 (3+4)	4	100
MDS372	Machine Learning	90	8	5	150
	ELECTIVE (Statistics - Concepts Based)				
MDS332A	Categorical Data Analysis	45	4	3	100
MDS 332B	Multivariate Analysis				
MDS332C	Stochastic Processes				
MDS381	SEMINAR	30	3	2	50
VAC1	Cloud Services	30	3	2	100
HED	HOLISTIC EDUCATION		1	1	50
	Total		**30**	**20**	**650**

TRIMESTER-IV

Course Code	Course Title	Course hrs	Hours Per Week	Credits	Marks
MDS431	Data driven Modelling and Visualization	30	3	2	100
MDS432	Time Series and Forecasting Techniques	60	5	4	100
MDS471	Neural Networks and Deep Learning	90	8	5	150
	ELECTIVES (Data Science)				
MDS472A	Web Analytics	60 (3+2)	5	3	100
MDS472B	IoT Analytics				
MDS472C	Natural Language Processing				
MDS473D	Image and Video Analytics				
MDS481	PROJECT-I (Web project with Data Science concepts)	60	5	2	100
MDS482	RESEARCH PROBLEM identification	30	3	1	50
	Total		**30**	**17**	**600**

TRIMESTER-V

Course Code	Course Title	Course hrs	Hours Per Week		Marks
MDS571	Big Data Analytics	90	8	5	150
	ELECTIVE - 1 (Applied Statistics)				
MDS531A	Econometrics	60	5	4	100
MDS531B	Bayesian Inference				
MDS531C	Bio-statistics				
	ELECTIVE-2 (Emerging analysis paradigms)				
MDS572A	Evolutionary Algorithms	60	5	4	100
MDS572B	Quantum Machine Learning				
MDS572C	Reinforcement Learning				
	ELECTIVE-3 (Unconventional Data Analysis)				
MDS573A	Geospatial Data Analytics	60	5	4	100
MDS573B	Bio-Informatics				

MDS573C	Graph Analytics				
MDS581	Project - II (Research Project/ Data Science Capstone Project)	60	5	2	100
	Total		**27**	**18**	**550**

TRIMESTER-VI

Course Code	Course Title	Course hrs	Hours 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: 5L-0T-0P

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	LRNG Needs
CO1	Understand the essence of research and the importance of research methods and methodology	**National**
CO2	Explore the fundamental concepts of data science	**Global**
CO3	Understand various machine learning algorithms used in data science process	**Global**
CO4	Learn to think through the ethics surrounding privacy, data sharing and algorithmic decision-making	**National**
CO5	Create scientific reports according to specified standards	**Global**

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					**Yes**

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2		1		
CO2	2	3			1	
CO3	2	3			1	
CO4		2		3		1
CO5	2				3	1

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CIA1 (20 MARKS)	CIA2 (50 MARKS)	CIA3 (20 MARKS)	ESE (100 MARKS)
CO1	10			20
CO2	10	25		20
CO3		25	05	20
CO4			05	20
CO5			10	20

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	CO'S MAPPED
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:	CO1

	Different Research Designs, Basic Principles of Experimental Designs, Developing a Research Plan. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** C. R. Kothari, *Research Methodology Methods and Techniques*. 3rd. ed. New Delhi: New Age International Publishers, Reprint 2014. Zina O'Leary, *The Essential Guide of Doing Research*. New Delhi: PHI, 2005.	
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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Davy Cielen and Arno Meysman, *Introducing Data Science*. Simon and Schuster, 2016.	CO2
UNIT 3 **Teaching Hours:12**	**Machine Learning** Machine learning – Modeling Process – Training model – Validating model – Predicting new observations – Supervised learning algorithms – Unsupervised learning algorithms.	CO3, CO4

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Davy Cielen and Arno Meysman, *Introducing Data Science*. Simon and Schuster, 2016.	
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.** **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Class Activity **Essential Reading:** M. Loukides, H. Mason, and D. Patil, *Ethics and Data Science*. O'Reilly Media, 2018. Zina O'Leary, *The Essential Guide of Doing Research*. New Delhi: PHI, 2005.	CO3, CO5
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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Class Activity **Essential Reading:** C. R. Kothari, *Research Methodology Methods and Techniques*. 3rd. ed. New Delhi: New Age International Publishers, Reprint 2014.	CO1, CO4

	Zina O'Leary, *The Essential Guide of Doing Research*. New Delhi: PHI, 2005.	

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: New Age 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] Sinan Ozdemir, *Principles of Data Science learn the techniques and math you need to start making sense of your data*. Birmingham Packt December, 2016.
[5] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* 4thed. SAGE Publications, 2014.
[6] Kumar, *Research Methodology: A Step-by-Step Guide for Beginners*. 3rd. ed. Indian: PE, 2010.

MDS 132: PROBABILITY AND DISTRIBUTION THEORY

Total Teaching Hours for Trimester: 60

No of hours per week: 5L-0T-0P

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	LRNG Needs
CO1	Describe random event and probability of events	**Global**
CO2	Identify various discrete and continuous distributions and their usage	**Global**
CO3	Evaluate condition probabilities and conditional expectations	Regional
CO4	Apply Chebychevs inequality to verify the convergence of sequence in probability	**National**

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					**Yes**

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	2				

CO2	1	2				2
CO3	2				1	1
CO4	2		3		1	

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	**CIA1 (20 MARKS)**	**CIA2 (50 MARKS)**	**CIA3 (20 MARKS)**	**ESE (100 MARKS)**
CO1	10	10		25
CO2	10	20		25
CO3		20	10	25
CO4			10	25

CO-UNIT MAPPING:

UNIT	**TOPICS/ SUB TOPICS**	**CO'S MAPPED**
UNIT 1 **Teaching Hours:12**	**Descriptive Statistics and Probability** Data – types of variables: numeric vs categorical - measures of central tendency – measures of dispersion - random experiment - sample space and random events – probability - probability axioms - finite sample space with equally likely outcomes - conditional probability - independent events - Baye's theorem **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Introduction to probability models. Ross, Sheldon M. 12th Edition, Academic Press, 2019.	CO1 ,CO3
UNIT 2 **Teaching Hours:12**	**Probability Distributions for Discrete Data** Random variable – data as observed values of a random variable - expectation – moments & moment generating	CO1 ,CO2

	function - mean and variance in terms of moments - discrete sample space and discrete random variable – Bernoulli experiment and Binary variable: Bernoulli and binomial distributions – Count data: Poisson distribution – over dispersion in count data: negative binomial distribution – dependent Bernoulli trails: hypergeometric distribution (mean and variances in terms of mgf). **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Fundamentals of Applied Mathematics, S.C. Gupta and V.K. Kapoor (New Edition)	
UNIT 3 **Teaching Hours:12**	**Probability Distributions For Continuous Data** Continuous sample space - Interval data - continuous random variable – uniform distribution - normal distribution (Gaussian distribution) – modeling lifetime data: exponential distribution, gamma distribution, Weibull distribution (Applications in Data science). **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Fundamentals of Applied Mathematics, S.C. Gupta and V.K. Kapoor (New Edition)	CO1, CO2
UNIT 4 **Teaching Hours:12**	**Jointly Distributed Random Variables** Joint distribution of vector random variables – joint moments – covariance – correlation - independent random variables - conditional distribution – conditional expectation - sampling distributions: chi-square, t, F (pdf's & properties). **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity.	CO1,CO3

	Essential Reading: Introduction to the theory of statistics. A.M Mood, F.A Graybill and D.C Boes, Tata McGraw-Hill, 3rd Edition (Reprint), 2017.	
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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Fundamentals of Applied Mathematics, S.C. Gupta and V.K. Kapoor (New Edition)	CO4

Essential Reading

[1] Introduction to the theory of statistics. A.M Mood, F.A Graybill and D.C Boes, Tata McGraw-Hill, 3rd Edition (Reprint), 2017.
[2] Introduction to probability models. Ross, Sheldon M. 12th Edition, Academic Press, 2019.
[3] Fundamentals of Applied Mathematics, S.C. Gupta and V.K. Kapoor (New Edition)

Recommended Reading

[1] A first course in probability. Ross, Sheldon, 10th Edition. Pearson, 2019.
[2] An Introduction to Probability and Statistics. V.K Rohatgi and Saleh, 3rd Edition, 2015

MDS133: MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE - I

Total Teaching Hours for Trimester: 45

No of hours per week: 4L-0-0P

Max Marks: 100 **Credits: 3**

Course Type: Major

Course Description

Linear Algebra plays a fundamental role in the theory of Data Science. This course aims at introducing the basic notions of vector spaces and its spans and orthogonalization, linear transformation and the use of its matrix bijections in applications to Data Science.

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

No.	Course Outcomes	LRNG Needs
CO1	Understand the properties of Vector spaces	**Global**
CO2	Use the properties of Linear Maps in solving problems on Linear Algebra	**Global**
CO3	Demonstrate proficiency on the topics Eigenvalues, Eigenvectors and Inner Product Spaces	
CO4	Apply mathematics for some applications in Data Science	**Global**

Cross Cutting Issues:

Employability	Skill Development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**				**YES**	

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6

CO1	3	3	3	3	3	3
CO2	3	3	2	3	1	2
CO3	3	3	3	3	3	3
CO4	3	3	3	3	3	3

CO-ASSESSMENT MAPPING-THEORY COMPONENT:

Course Outcomes /Unit	**CIA I (20 MARKS)**	**CIA II (50 MARKS)**	**CIA III (20 MARKS)**	**ESE (100 MARKS)**
CO1	10	20		17.50
CO2	10	20		17.50
CO3		10	20	17.50
CO4				47.50

CO-UNIT MAPPING:

UNIT	**TOPICS/ SUB TOPICS**	***CO'S MAPPED**
UNIT 1 **Teaching Hours: 09L**	**INTRODUCTION TO VECTOR SPACES** Vector Spaces: Definition and properties, Subspaces, Sums of Subspaces, Null space , Column space, Direct Sums, Span and Linear Independence, Bases, dimension, rank. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. David C. Lay, Steven R. Lay, Judi J. McDonald (2016) Linear algebra and its applications. Pearson. 2. S. Axler, Linear algebra done right, Springer, 2017. 2. Strang, G. (2006) Linear Algebra and its Applications: Thomson Brooks. Cole, Belmont, CA, USA.	**CO1, CO2**

UNIT 2 **Teaching Hours:09L**	**LINEAR TRANSFORMATIONS** Algebra of Linear Transformations, Null spaces and Injectivity, Range and Surjectivity, Fundamental Theorems of Linear Maps- Cayley-Hamilton theorem - Orthonormal basis. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. David C. Lay, Steven R. Lay, Judi J. McDonald (2016) Linear algebra and its applications. Pearson. 2. S. Axler, Linear algebra done right, Springer, 2017. 2. Strang, G. (2006) Linear Algebra and its Applications: Thomson Brooks. Cole, Belmont, CA, USA.	**CO1, CO2**
UNIT 3 **Teaching Hours: 09L**	**EIGENVALUES AND EIGENVECTORS** Invariant Subspaces, Polynomials applied to Operators – Upper-Triangular matrices, Diagonal matrices, Invariant Subspaces on real vector Spaces Eigen values and Eigen vectors – Characteristic equation – Diagonalization - Upper Triangular matrices - Invariant Subspaces on Real Vector Spaces **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. David C. Lay, Steven R. Lay, Judi J. McDonald (2016) Linear algebra and its applications. Pearson. 2. S. Axler, Linear algebra done right, Springer, 2017. 2. Strang, G. (2006) Linear Algebra and its Applications: Thomson Brooks. Cole, Belmont, CA, USA.	**CO2,CO3**

UNIT 4 **Teaching Hours: 09L**	**INNER PRODUCT SPACES** Inner Products and Norms – Orthogonality - Orthogonal Bases – Orthogonal Projections –Gram-Schmidt process - Least square problems – Applications to Linear models **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. David C. Lay, Steven R. Lay, Judi J. McDonald (2016) Linear algebra and its applications. Pearson. 2. S. Axler, Linear algebra done right, Springer, 2017. 2. Strang, G. (2006) Linear Algebra and its Applications: Thomson Brooks. Cole, Belmont, CA, USA.	**CO2,CO3**
UNIT 5 **Teaching Hours: 09L**	**BASIC MATRIX METHODS FOR APPLICATIONS** Matrix Norms –Singular value decomposition-Householder Transformation and QR decomposition-Non Negative Matrix Factorization – bidiagonalization **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. David C. Lay, Steven R. Lay, Judi J. McDonald (2016) Linear algebra and its applications. Pearson. 2. S. Axler, Linear algebra done right, Springer, 2017. 2. Strang, G. (2006) Linear Algebra and its Applications: Thomson Brooks. Cole, Belmont, CA, USA.	**CO4**

Essential References

[1] David C. Lay, Steven R. Lay, Judi J. McDonald (2016) Linear algebra and its applications. Pearson. 2. S. Axler, Linear algebra done right, Springer, 2017.

[2] Strang, G. (2006) Linear Algebra and its Applications: Thomson Brooks. Cole, Belmont, CA, USA.

Recommended References

[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: 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	LRNG Needs
CO1	Understand the fundamentals of programming languages.	**National**
CO2	Understand the design paradigms of programming languages.	**Global**
CO3	To examine expressions, subprograms and their parameters.	**Global**

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6
CO1	2		--		--	2

CO2		2	--	2	--	1
CO3		3	--	2	--	2

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	**CIA1 (25 MARKS)**	**CIA2 (25 MARKS)**	**ESE (50 MARKS)**
CO1	15	5	15
CO2	10	10	15
CO3		10	20

CO-UNIT MAPPING:

UNIT	**TOPICS/ SUB TOPICS**	**CO'S MAPPED**
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). **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Linux: The Complete Reference, sixth edition, Richard Petersen,2017	CO1 ,CO2
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).	CO1 ,CO2

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Linux: The Complete Reference, sixth edition, Richard Petersen,2017	
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).	CO1 ,CO2, CO3

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 for Semester: 30

No of hours per week: 3L-0-0P

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	LRNG Needs
CO1	Demonstrate, present and visualize data in various forms, statistically.	**Global**
CO2	Understand and apply descriptive statistics.	**Global**
CO3	Evaluation of probabilities for various kinds of random events.	**Global**

Cross Cutting Issues:

Employability	Skill Development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**				**YES**	

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	3	3	3	3	3
CO2	3	3	2	3	3	3
CO3	3	3	3	3	3	3

CO-ASSESSMENT MAPPING-THEORY COMPONENT:

Course Outcomes /Unit	CIA I (25 MARKS)	CIA II (25 MARKS)	ESE (50 MARKS)
CO1	25		10
CO2		25	15
CO3			25

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	*CO'S MAPPED
UNIT 1 **Teaching Hours: 08L**	**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). **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** .1. Gupta S.C and Kapoor V.K, *Fundamentals of Mathematical Statistics*, 12th edition, Sultan Chand & Sons, New Delhi, 2020.	**CO1**
UNIT 2 **Teaching Hours:06L**	**DESCRIPTIVE STATISTICS I** Measures of location or central tendency: Arithmetic mean - Median - Mode - Geometric mean - Harmonic mean.	**CO2**

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Gupta S.C and Kapoor V.K, *Fundamentals of Mathematical Statistics*, 12th edition, Sultan Chand & Sons, New Delhi, 2020.	
UNIT 3 **Teaching Hours: 06L**	**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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** .1. Gupta S.C and Kapoor V.K, *Fundamentals of Mathematical Statistics*, 12th edition, Sultan Chand & Sons, New Delhi, 2020.	**CO2**
UNIT 4 **Teaching Hours: 10L**	**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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** .1. Gupta S.C and Kapoor V.K, *Fundamentals of Mathematical Statistics*, 12th edition, Sultan Chand & Sons, New Delhi, 2020.	**CO3**

Essential References

[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 References

[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 for Semester: 30

No of hours per week: 3L-0-0P

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	LRNG Needs
CO1	Demonstrate the Basic file, directory commands	**National**
CO2	Understand the Unix system environment	**Global**
CO3	Apply shell programming concepts to solve given problem	**Global**

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2		--		--	2
CO2		2	--	2	--	1
CO3		3	--	2	--	2

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CIA1 (25 MARKS)	CIA2 (25 MARKS)	ESE (50 MARKS)
CO1	15	5	15
CO2	10	10	15
CO3		10	20

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	CO'S MAPPED
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 **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Linux: The Complete Reference, sixth edition, Richard Petersen, 2017	CO1 ,CO2
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 **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity	CO1 ,CO2

	Essential Reading: Linux: The Complete Reference, sixth edition, Richard Petersen, 2017	
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 s expr command Shell Screen Interface, read and echo statement,Shell Scri Conditional Control Structures – if statement,Case statement,Looping C while,for,Jumping Control Structures – break, continue, exit.	CO1 ,CO2, CO3

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 Semester: 90

No of hours per week: 4L-0T-4P

Max Marks: 150 **Credits: 5**

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 completion of the course students will be able to

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

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**	**Yes**				**Yes**

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	-	-	1	-	-
CO2	-	2	2	2	-	2
CO3	-	2	-	3	-	3

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CAT1 1	CAT2	CAT3	CAC1	CAC 2	Regular Program evaluations	ATTD 8 marks
CO1	9	7	10	5		8	Not applicable
CO2	9	8	10	5	10	13	
CO3		8	10	5	12	13	

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	CO'S MAPPED
UNIT 1 **Teaching Hours:18**	**Introduction** **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 Exercises** 1. Demonstrate usage of branching and looping statements 2. Demonstrate Recursive functions 3. Demonstrate Lists **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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	CO1

UNIT 2 **Teaching Hours:18**	**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 and Dictionaries 2. Demonstrate inheritance and exception handling 3. Demonstrate use of "re" **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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	CO1 ,CO2
UNIT 3 **Teaching Hours:18**	**USING NUMPY** Basics of NumPy - Computation on NumPy - Aggregations - Computation on Arrays- Comparisons, Masks and Boolean Arrays - Fancy Indexing-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 **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity	CO1 ,CO2,CO3

	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	
UNIT 4 **Teaching Hours:18**	**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() **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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.	CO1 ,CO2,CO3
UNIT 5 **Teaching Hours:18**	**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 - Three- Dimensional Plotting in Matplotlib. **Lab Exercises** 1. Demonstrate Line plot and Scatter plat 2. Demonstrate 3D plotting	CO2,CO3

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **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	

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: 03P

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	LRNG Needs
CO1	Demonstrate the data management using excel features.	**National**
CO2	Analyze the given problem and solve using Excel.	**Global**
CO3	Infer the building blocks of excel, excel shortcuts, sample data creation	**Global**

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2		--		--	2
CO2	1	2	--	2	--	1
CO3	1	3	--	2	--	2

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CIA1 (15 MARKS)	CIA2 (15 MARKS)	Regular Lab Programs (20 Marks)	ESE (50 MARKS)
CO1	**10**	**5**	**8**	**15**
CO2	**5**	**5**	**7**	**15**
CO3		**5**	**5**	**20**

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	CO'S MAPPED
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	CO1 ,CO2

	3. Data Collection 4. Excel Charts	
UNIT 2 **Teaching Hours :10**	**Files** **Teaching Hours: 5** 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 5.Import data 6.Export data 7.Creating database 8.Data tools	CO1 ,CO2
UNIT 3 **Teaching Hours :10**	**Unit-III** **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. 9.Excel functions.	CO1 ,CO2, CO3

Essential Reading:

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

Recommended Reading

[2] Paul M, Microsoft Excel 2019 formulas and functions, Pearson Eduction, 2019

MDS231: Design and Analysis of Algorithms

Total Teaching Hours for Semester: 45

No of hours per week: 3L-0T-0P

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	LRNG Needs
CO1	Analyze asymptotic and absolute runtime and memory demands of algorithms	Global
CO2	Apply classical sorting, searching, optimization and graph algorithms.	Global
CO3	Understand basic techniques for designing algorithms, including the techniques of recursion, divide-and-conquer, greedy algorithm etc.	Global
CO4	Evaluate algorithm efficiency mathematically	Global

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					**Yes**

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	3	3	1	2	2
CO2	3	3	3	1	2	2
CO3	3	3	2	1	1	2
CO4	3	3	2	1	2	2

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CIA1 (20 MARKS)	CIA 2 (50 MARKS)	CIA3 (20 MARKS)	ESE (100 MARKS)
CO1	10	20		20
CO2	10	20		30
CO3		10	10	30
CO4			10	20

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	CO'S MAPPED
UNIT 1 **Teaching Hours:9**	**Introduction** Algorithms, Analyzing algorithms, Complexity of algorithms, Growth of functions, Performance measurements, Sorting and order Statistics - Shell sort, Heap sort, Sorting in linear time. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** 1. Coreman, Rivest, Lisserson, "An Introduction	CO1, CO2

	to Algorithm", PHI, 2001 2. Horowitz & SAHANI," Fundamental of computer Algoritm", Galgotia Publications, 2nd Edition.	
UNIT 2 **Teaching Hours:9**	**Advanced Data Structures** Red-Black trees, B – trees, Binomial Heaps, Fibonacci Heaps, Tries, skip list. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** 3. Coreman, Rivest, Lisserson, "An Introduction to Algorithm", PHI, 2001 4. Horowitz & SAHANI," Fundamental of computer Algoritm", Galgotia Publications, 2nd Edition.	CO1 ,CO2
UNIT 3 **Teaching Hours:9**	**Divide and Conquer** Quick sort, Merge sort, Finding maximum and minimum,Matrix Multiplication, 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 and Bellman Ford algorithms.Optimal merge patterns. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** 1. Coreman, Rivest, Lisserson, "An Introduction to Algorithm", PHI, 2001 2. Horowitz & SAHANI," Fundamental of computer Algoritm", Galgotia Publications, 2nd Edition.	CO1 ,CO2,CO3
UNIT 4 **Teaching Hours:9**	**Dynamic programming** with examples such as Knapsack, All pair shortest paths – Warshal's and Floyd's algorithms, Resource allocation problem. Backtracking, Branch and Bound with examples such	CO3,CO4

	as Travelling Salesman Problem, Graph Coloring, n-Queen Problem, Hamiltonian Cycles and Sum of subsets. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** 1. Coreman, Rivest, Lisserson, "An Introduction to Algorithm", PHI, 2001 2. Horowitz & SAHANI," Fundamental of computer Algoritm", Galgotia Publications, 2nd Edition.	
UNIT 5 **Teaching Hours:9**	**Selected Topics:** Algebraic Computation, Fast Fourier Transform, String Matching, Theory of NP-completeness, Approximation algorithms and Randomized algorithms. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** 1. Coreman, Rivest, Lisserson, "An Introduction to Algorithm", PHI, 2001 2. Horowitz & SAHANI," Fundamental of computer Algoritm", Galgotia Publications, 2nd Edition.	CO1 ,CO2,CO3,CO4

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.

MDS232 - MATHEMATICAL FOUNDATION FOR DATA SCIENCE - II

Total Teaching Hours for Semester: 45

No of hours per week: 3L-0T-0P

Max Marks: 100 **Credits: 3**

Course Type: Major

Course Description

This course aims at introducing data science related essential mathematics concepts such as fundamentals of topics on Calculus of several variables, Orthogonality, Convex optimization, and Graph Theory.

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

No.	Course Outcomes	LRNG Needs
CO1	Demonstrate the properties of multivariate calculus	Global
CO2	Use the idea of orthogonality and projections effectively	Global
CO3	Have a clear understanding of Convex Optimization	Global
CO4	Know the about the basic terminologies and properties in Graph Theory	Global

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	3	1	1	-	1
CO2	3	3	1	1	-	2
CO3	3	3	1	1	-	2
CO4	1	2	2	1	1	2

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CIA1 (20 MARKS)	CIA 2 (50 MARKS)	CIA3 (20 MARKS)	ESE (100 MARKS)
CO1	10	20		20
CO2	10	20		30
CO3		10	10	30
CO4			10	20

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	CO'S MAPPED
UNIT 1 **Teaching Hours :9**	**Calculus of Several Variables** Functions of Several Variables: Functions of two, three variables Limits and continuity in Higher Dimensions: Limits for functions o two variables, Functions of more than two variables - Partia Derivatives: partial derivative of functions of two variables, partia derivatives of functions of more than two variables - The Chain Rule chain rule on functions of two, three variables, chain rule on function defined on surfaces	CO1, CO2

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential References** M. D. Weir, J. Hass, and G. B. Thomas, Thomas' calculus. Pearson, 2016.	
UNIT 2 **Teaching Hours :9**	**Orthogonality** Perpendicular vectors and Orthogonality - Inner Products and Projections onto lines - Projections of Rank one - Projections and Least Squares Approximations - Projection Matrices - Orthogonal Bases, Orthogonal Matrices. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential References:** G Strang, Linear Algebra and its Applications, 4th ed., Cengage , 2006.	CO1, CO2
UNIT 3 **Teaching Hours :9**	**Introduction to Convex Optimization** Affine and Convex Sets: Lines and Line segments, affine sets, affin dimension andrelative interior, convexsets, cones - Hyperplanes and half-spaces - Euclidean balls and ellipsoids- Norm balls and Norm cones – polyhedral. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** S. P. Boyd and L.Vandenberghe, Convex optimization.Cambridge Univ. Pr., 2011.	CO1,CO2, CO3
UNIT 4 **Teaching**	**Graph Theory - Basics**	CO1, CO2, CO3

Hours :9	Graph Classes: Definition of a Graph and Graph terminology, isomorphism of graphs, Completegraphs, bipartite graphs, complete bipartite graphs-Vertex degree: adjacency and incidence, regular graphs - subgraphs, spanning subgraphs, induced subgraphs, removing or adding edges of a graph, removing vertices from graphs. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** J Clark, D A Holton, A first look at Graph Theory, Allied Publishers India, 1995.	
UNIT 5 **Teaching Hours :12**	**Graph Theory - More concepts** Matrix Representation of Graphs, Adjacency matrices, Incidence Matrices, Trees and its properties, Bridges (cut-edges), spanning trees, weighted Graphs, minimal spanning tree problems, Shortest path problems - Applications of Graph Theory **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** J Clark, D A Holton, A first look at Graph Theory, Allied Publishers India, 1995.	CO1, CO2, CO3, CO4

Essential References:

[1] M. D. Weir, J. Hass, and G. B. Thomas, Thomas' calculus. Pearson, 2016.

[2] G Strang, Linear Algebra and its Applications, 4th ed., Cengage, 2006.

[3] S. P. Boyd and L.Vandenberghe, Convex optimization.Cambridge Univ. Pr., 2011.

[4] J Clark, D A Holton, A first look at Graph Theory, Allied Publishers India, 1995.

Recommended References:

[1] J. Patterson and A. Gibson, Deep learning: a practitioner's approach. O'Reilly Media, 2017.

[2] S. Sra, S. Nowozin, and S. J. Wright, Optimization for machine learning. MIT Press, 2012.

[3] D. Jungnickel, Graphs, networks and algorithms. Springer, 2014.

[4] D Samovici, Mathematical Analysis for Machine Learning and Data Mining, World Scientific Publishing Co. Pte. Ltd, 2018

[5] P. N. Klein, Coding the matrix: linear algebra through applications to computer science. Newtonian Press, 2015.

[6] K H Rosen, Discrete Mathematics and its applications, 7th ed., McGraw Hill, 2016

MDS271: Database Technologies

Total Teaching Hours for Semester: 75

No of hours per week: 3L-0T-4P

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

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	3	3	1	2	3
CO2	3	3	1	1	2	3
CO3	3	3	3	1	3	3

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CAT1 1	CAT2	CAT3	CAC1	CAC 2	Regular Program evaluatio ns	ATTD 5 marks
CO1	6	5				5	Not applicable
CO2	7	5	5	5	5	5	
CO3		5	7	5	5	6	
CO4			8		5	6	

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	CO'S MAPPED
UNIT 1 **Teaching Hours:15**	**Introduction** 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, Weak Entity Sets, Extended E-R features. **Lab Exercises** 1. Data Definition, 2. Table Creation	CO1

	3. Constraints **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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	
UNIT 2 **Teaching Hours:12**	**Relational model and database design** 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, Boyce-Codd Normal Form. **Lab Exercises** 1. Insert, Select, Update & Delete Commands 2. Nested Queries & Join Queries 3. Views **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** 1. Henry F. Korth and Silberschatz Abraham, "Database System Concepts", Mc.Graw Hill.	CO1 ,CO2

	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	
UNIT 3 **Teaching Hours:15**	**Data warehouse: the building blocks** 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. Importing source data structures 2. Design Target Data Structures 3. Create target multidimensional cube **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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	CO1 ,CO2,CO3
UNIT 4 **Teaching Hours:12**	**Data Integration and Data Flow (ETL)** Requirements, ETL Data Structures, Extracting, Cleaning and Conforming, Delivering Dimension Tables, Delivering Fact Tables, Real-Time ETL Systems	CO1 ,CO2,CO3

	Lab Exercises 1. Perform the ETL process and transform into data map 2. Create the cube and process it 3. Generating Reports 4. Creating the Pivot table and pivot chart using some existing data **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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	
UNIT 5 **Teaching Hours:12**	**NOSQL Databases** Introduction to NOSQL Systems, The CAP Theorem, Document-Based NOSQL Systems and MongoDB, NOSQL Key-Value Stores, Column-Based or Wide Column NOSQL Systems, Graph databases, Multimedia databases. **Lab Exercises** 1. MongoDB Exercise - 1 2. MongoDB Exercise - 2 **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** 1. Henry F. Korth and Silberschatz Abraham, "Database System Concepts", Mc.Graw Hill. 2. Thomas Cannolly and Carolyn	CO1 ,CO2,CO3

	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.	

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

Recommended Reading

[1] LiorRokach and OdedMaimon, Data Mining and Knowledge Discovery Handbook, Springer, 2nd edition, 2010.

MDS272: INFERENTIAL STATISTICS

Total Teaching Hours for Semester: 75

No of hours per week: 3L-0-4P

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	LRNG Needs
CO1	Demonstrate the concepts of population and samples	Global
CO2	Apply the idea of sampling distribution of different statistics in testing of hypothesis	Global
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.	Global

***Cross Cutting Issues:**

Employability	Skill Development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**				**Yes**	

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	-	1	3	3
CO2	-	-	-	2	3	3
CO3	3	3	2	3	3	3
CO4	3	3	2	3	3	3

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CAT1 1	CAT2	CAT3	CAC1	CAC 2	Regular Program evaluatio ns	ATTD 5 marks
CO1	6	5				5	Not applicable
CO2	7	5	5	5	5	5	
CO3		5	7	5	5	6	
CO4			8		5	6	

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	*CO'S MAPPED
UNIT 1 **Teaching Hours: 15**	**INTRODUCTION:** Population and Statistics – Finite and Infinite population – Parameter and Statistics – Types of sampling - Sampling Distribution – Sampling Error - Standard Error – Test of significance –concept of hypothesis – types of hypothesis – Errors in hypothesis- testing – Critical region – level of significance - Power of the test – p-value.	**CO1, CO2**

	1. Calculation of sampling error and standard error 2. Calculation of probability of critical region using standard distributions 3. Calculation of power of the test using standard distributions. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 12th edition, Sultan Chand & Sons, New Delhi, 2020. 2. Brian Caffo, Statistical Inference for Data Science, Learnpub, 2016.	
UNIT 2 **Teaching Hours:15**	**Testing of Hypothesis I** Concept of large and small samples – Tests concerning a single population mean for known σ (and unknown σ) – equality of two means for known σ (and unknown σ) – Test for Single variance - Test for equality of two variance for normal population – Tests for single proportion – Tests of equality of two proportions for the normal population. Lab Exercises 1. Test of the single sample mean for known and unknown σ. 2. Test of equality of two means when known and unknown σ. 3. 6. Tests of single variance and equality of variance for large samples. 4. 7. Tests for single proportion and equality of two proportion for large samples. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity	**CO3**

	Essential Reading 1. Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 12th edition, Sultan Chand & Sons, New Delhi, 2020. 2. Brian Caffo, Statistical Inference for Data Science, Learnpub, 2016.	
UNIT 3 **Teaching Hours: 15**	**Testing of Hypothesis II** Students t-distribution and its properties (without proofs) – Single sample mean test – Independent sample mean test – Paired sample mean test – Tests of proportion (based on t distribution) – F distribution and its properties (without proofs) – Tests of equality of two variances using F-test – Chi-square distribution and its properties (without proofs) – chisquare test for independence of attributes – Chi-square test for goodness of fit. **Lab Exercises** 1. Single sample mean test 2. Independent and Paired sample mean test 3. Tests of proportion of one and two samples based on t-distribution **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 12th edition, Sultan Chand & Sons, New Delhi, 2020. 2. Brian Caffo, Statistical Inference for Data Science, Learnpub, 2016.	**CO3**
UNIT 4 **Teaching Hours: 15**	**Analysis of Variance** Meaning and assumptions - Fixed, random and mixed effect models - Analysis of variance of one-way and two-way classified data with and without interaction	**CO3**

	effects – Multiple comparison tests: Tukey's method - critical difference. **Lab Exercises** 1. Test of equality of two variances 2. Chi-square test for independence of attributes and goodness of fit. 3. Construction of one-way ANOVA 4. Construction of two-way ANOVA with interaction 5. Construction of two-way ANOVA without interaction **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 12th edition, Sultan Chand & Sons, New Delhi, 2020. 2. Brian Caffo, Statistical Inference for Data Science, Learnpub, 2016.	
UNIT 5 **Teaching Hours:15**	**Nonparametric Tests** 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. Multiple comparision test using Tukey's method and critical difference methods 2. Test of one sample using Run and sign tests 3. Test of paried sample using Wilcoxon signed rank test 4. Test of two samples using Run test and Median test	**CO4**

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 12th edition, Sultan Chand & Sons, New Delhi, 2020. 2. Brian Caffo, Statistical Inference for Data Science, Learnpub, 2016.	

Essential References

[1] Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 12th edition, Sultan Chand & Sons, New Delhi, 2020.

[2] Brian Caffo, Statistical Inference for Data Science, Learnpub, 2016.

Recommended References

[1] Walpole R.E, Myers R.H and Myers S.L, Probability and Statistics for Engineers and Scientists, 9th edition, Pearson, New Delhi, 2017.

[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 (P) Ltd, New Delhi, 2012.

[4] Rohatgi V.K and Saleh E, An Introduction to Probability and Statistics, 3rd edition, JohnWiley & Sons Inc, New Jersey, 2015.

MDS273: FULL STACK WEB DEVELOPMENT

Total Teaching Hours for Semester: 75

No of hours per week: 3L-0-4P

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	LRNG Needs
CO1	Apply JavaScript, HTML5, and CSS3 effectively to create interactive and dynamic websites	**National**
CO2	Describe the main technologies and methods currently used in creating advanced web applications	**National**
CO3	Design websites using appropriate security principles, focusing specifically on the vulnerabilities inherent in common web implementations	**Local**
CO4	Create modern web applications using MEAN	**Global**

***Cross Cutting Issues:**

Employability	Skill Development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**				**Yes**	

***CO-PO MAPPING:**

Course Outcomes /Programme Outcomes	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**
CO1	2	-	-	1	2	2
CO2	3	-	-	-	-	2
CO3	2	-	-	3	3	2
CO4	3	2	2	-	-	2

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	**CAT1 1**	**CAT2**	**CAT3**	**CAC1**	**CAC 2**	**Regular Program evaluatio ns**	**ATTD 5 marks**
CO1	6	5				5	Not applicable
CO2	7	5	5	5	5	5	
CO3		5	7	5	5	6	
CO4			8		5	6	

CO-UNIT MAPPING:

UNIT	**TOPICS/ SUB TOPICS**	***CO'S MAPPED**
UNIT 1 **Teaching Hours: 10L**	**OVERVIEW OF WEB TECHNOLOGIES AND HTML5** 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**	CO1,CO2

	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 **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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.	
UNIT 2 **Teaching Hours:10L**	**XML AND AJAX** XML-Documents and Vocabularies-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 **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** [1] Internet and World Wide Web:How to Program, Paul Deitel , Harvey Deitel & Abbey Deitel, Pearson Education, 5th Edition, 2018.	CO1,CO2

	[2] HTML 5 Black Book (Covers CSS3, JavaScript, XML, XHTML, AJAX, PHP, jQuery), DT Editorial Services, Dreamtech Press, 2nd Edition, 2016.	
UNIT 3 **Teaching Hours: 10L**	**CLIENT SIDE SCRIPTING** 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 Event Handling 2. Create a web application using AngularJS with Forms **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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.	CO1,CO2,CO3
UNIT 4 **Teaching Hours: 10L**	**SERVER SIDE SCRIPTING** Introduction to Node.js-REPL Terminal-Package Manager(NPM)-Node.js Modules and filesystem-Node.js Events-Debugging Node JS Application-File System and streams-Testing Node JS with jasmine **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity	CO1,CO2,CO3

	Lab Exercises 1. Implement a single page web application using Angular JS CRUD Operation using AngularJS 2. Implement web application using AJAX with JSON 3. Demonstrate to fetch the information from an XML file with AJAX **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. Education, 2017.	
UNIT 5 **Teaching Hours: 10L**	**NODE JS WITH MYSQL** Introduction to MySQL- Performing basic database operation(DML) (Insert, Delete, Update, Select)- Prepared Statement- Uploading Image or File to MySQL- Retrieve Image or File from MySQL **Lab Exercises** 1. Demonstrate Node.js file system module 2. Implement Mysql with Node.JS 3. Implement CRUD Operation using MongoDB **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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.	CO1,CO2,CO3,CO4

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.

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

MDS331: REGRESSION MODELING

Total Teaching Hours for Semester: 45

No of hours per week: 4L-0-0P

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	LRNG Needs
CO1	Formulate the linear regression model and its application to real data.	**Global**
CO2	Understand and identify the various assumptions of linear regression models.	**Global**
CO3	Identify the correct model using model selection and variable selection criteria.	**Global**
CO4	Ability to use and understand generalizations of the linear model to binary and count data.	**Global**

Cross Cutting Issues:

Employability	Skill Development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**				**YES**	

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	3	3	3	3	3
CO2	3	3	2	3	1	2
CO3	3	3	3	3	3	3
CO4	3	3	3	3	3	3

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CIA I (20 MARKS)	CIA II (50 MARKS)	CIA III (20 MARKS)	ESE (100 MARKS)
CO1	10	20		17.50
CO2	10	20		17.50
CO3		10	20	17.50
CO4				47.50

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	*CO'S MAPPED
UNIT 1 **Teaching Hours: 10L**	**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.	**CO1, CO2**

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **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.	
UNIT 2 **Teaching Hours:09L**	**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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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.	**CO1, CO2**
UNIT 3 **Teaching Hours: 09L**	**Model Adequacy** Residual analysis; Departures from underlying assumptions: Multicollinearity, Heteroscedasticity, Autocorrelation, Effect of outliers. Diagnostics and remedies.	**CO2,CO3**

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **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.	
UNIT 4 **Teaching Hours: 08L**	**Model Selection Criteria** Model selection criteria: R-Square, Adjusted R-Square, Mean Square error criteria; Variable selection criteria: Forward, Backward and Stepwise procedures. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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.	**CO2,CO3**
UNIT 5 **Teaching Hours: 09L**	**Non-Linear Regression** Introduction to nonlinear regression, Least squares in the nonlinear case and estimation of parameters, Models for binary and count response variable. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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.	**CO4**

	2. Chatterjee S and Hadi A, *Regression Analysis by Example*, 4th edition, John Wiley and Sons Inc, New York, 2015.	

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.

MDS371: JAVA PROGRAMMING

Total Teaching Hours for Semester: 75

No of hours per week: 3L-0-4P

Max Marks: 100 **Credits: 4**

Course Type: Major

Course Description

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

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

No.	Course Outcomes	LRNG Needs
CO1	Apply object-oriented programming structures in Java to solve real world problems	**National**
CO2	Demonstrate understanding of generics and collections framework	**Global**
CO3	Design programs for multi-threaded environment	**National**
CO4	Analyze and visualize data using various libraries	**Local**

***Cross Cutting Issues:**

Employability	Skill Development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**				**Yes**	

***CO-PO MAPPING:**

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	3	-	1	-	2

CO2	3	3	-	-	-	2
CO3	2	3	-	3	-	2
CO4	3	3	-	-	-	-

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	**CAT1 1**	**CAT2**	**CAT3**	**CAC1**	**CAC 2**	**Regular Program evaluations**	**ATTD 5 marks**
CO1	6	5				5	Not applicable
CO2	7	5	5	5	5	5	
CO3		5	7	5	5	6	
CO4			8		5	6	

CO-UNIT MAPPING:

UNIT	**TOPICS/ SUB TOPICS**	**CO'S MAPPED**
UNIT 1 **Teaching Hours: 15**	**INTRODUCTION: OVERVIEW OF JVM AND JAVA BASICS** Overview of JVM Introduction to JVM-JVM Architecture-JDK&JRE-Class Loader-Overview of Bootstrap, Extension and Application Class Loader Java Basics Class and Object Concept-Method Overloading and Overriding-Constructor-this and static keyword-finalize () method in java Inheritance in Java	CO1

	Inheritance Basics - Multilevel Hierarchy- Using super - Dynamic Method Dispatch- 1. Implement the concept of class, data members, member functions and access specifiers. 2. Implement the concept of function overloading & Constructor overloading 3. Implement the static keyword – static variable, static block, static function and static **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity ***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 Easy*Packt,2017.	
UNIT 2 **Teaching Hours:15**	**INTERFACES & PACKAGES AND EXCEPTION HANDLING IN JAVA** Abstract keyword- Using final with inheritance – Aggregation and Composition in Java Interfaces and Packages Defining Interfaces - Implementing Interfaces - Extending Interfaces- Creating Packages - Importing Packages - Interfaces in a Package. Exception Handling in Java try-catch-finally mechanism - throw statement - throws statement - Built-in-Exceptions – Custom Exceptions. **Lab Exercises** 1. Implement the concept of inheritance, super, abstract and final keywords 2. Implement package and interface 3. Implement Exception Handing in java **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity	CO1

	***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 Easy*Packt,2017.	
UNIT 3 **Teaching Hours: 15**	**EXEPTION HANDLING, MULTITHREADING and GENERICS** Thread Model - Life cycle of a Thread - Java Thread Priorities - Runnable interface and Thread Class- Thread Synchronization – Inter Thread Communication. Generics Generics Concept - General Form of a Generic Class – Bounded Types – Generic Class Hierarchy - Generic Interfaces – Restrictions in Generics 1. Implement multithreading – Thread class, Runnable interface, synchronization and thread communication. 2. Implementation of Generics Concepts **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity ***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 Easy*Packt,2017.	CO2
UNIT 5 **Teaching Hours:15**	**THE COLLECTIONS FRAMEWORK** The Collections Framework The Collections Overview – Collection Interface – List Interface – Set Interface – SortedSet Interface – Queue Interface - ArrayList Class – LinkedList Class	CO3

	– HashSet Class – Using an Iterator – The For Each Statement. 1. Implement collection Interfaces 2. Implementation of collections classes **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity ***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 Easy*Packt,2017.	
UNIT V: Teaching Hours:15	**Data Science in Java** Data Science Libraries- data processing library, Math and Stats libraries, machine learning and data mining libraries. Standard Java Library -Collections-Input-Output, Acessing Data-CSV, JSON, DataFrames. Working with mean median mode, Understanding plots and graphs **Lab Exercises:** 1. Lab exercise on handling CSV and JSON files 2. Lab exercise on data visualization **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 Easy*Packt,2017.	CO4

Essential References:

[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 Easy*Packt,2017.

Recommended References:

[1] Bloch, J. (2016). *Effective java*. Pearson Education India.

[2] Schildt, H., & Coward, D. (2014). *Java: the complete reference*. New York: McGraw-Hill Education.

MDS372 - MACHINE LEARNING

Total Teaching Hours for Semester: 90

No of hours per week: 4L-0T-4P

Max Marks: 150 **Credits: 5**

Course Type: Major

Course Description

The objective of this course is to provide an introduction to 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	LRNG Needs
CO1	Understand the basic principles of machine learning techniques.	**Global**
CO2	Understand how machine learning problems are formulated and solved	**National**
CO3	Apply machine learning algorithms to solve real world problems.	**Regional**

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**	**Yes**				**Yes**

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	3	-	1	-	1

CO2	3	-	-	-	2	1
CO3	-	3	-	2	1	1

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	**CAT1 [37]**	**CAT2 [45]**	**CAT3 [60]**	**CAC1 [30]**	**CAC 2 [45]**	**Regular Program evaluations [68]**	**ATTD [15]**
CO1	12	15	20	10	15	20	5
CO2	12	15	20	10	15	24	5
CO3	13	15	20	10	15	24	5

CO-UNIT MAPPING:

UNIT	**TOPICS/ SUB TOPICS**	**CO'S MAPPED**
UNIT 1 **Teaching Hours:18**	**Introduction** 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. Introduction to Parametric methods-Maximum Likelihood Estimation: Bernoulli Density-Multinomial Density-Gaussian Density, Nonparametric Density Estimation: Histogram Estimator-Kernel Estimator-K- Nearest Neighbour Estimator. **Lab Exercises** 1. Data Exploration using parametric methods 2. Data Exploration using non-parametric methods	CO1, CO3

	3. Regression analysis **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** E. Alpaydin, Introduction to Machine Learning, 3rd Edition, MIT Press, 2014.	
UNIT 2 **Teaching Hours:18**	**Dimensionality Reduction** Dimensionality Reduction: Introduction- Subset Selection-Principal Component Analysis, Feature Embedding-Factor Analysis-Singular Value Decomposition-Multidimensional Scaling-Linear Discriminant Analysis- Bayesian Decision Theory. **Lab Exercises** 1. Data reduction using Principal Component Analysis 2. Data reduction using multi-dimensional scaling **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** E. Alpaydin, Introduction to Machine Learning, 3rd Edition, MIT Press, 2014.	CO1
UNIT 3 **Teaching Hours:18**	**Supervised Learning - I** 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.	CO2, CO3

	Lab Exercises 1. Linear discrimination 2. Logistic discrimination 3. Classification using kernel machines **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** E. Alpaydin, Introduction to Machine Learning, 3rd Edition, MIT Press, 2014.	
UNIT 4 **Teaching Hours:18**	**Supervised Learning - II** 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 Exercises** 1. Classification using MLP 2. Ensemble Learning **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** E. Alpaydin, Introduction to Machine Learning, 3rd Edition, MIT Press, 2014.	CO2, CO3
UNIT 5 **Teaching Hours:18**	**Unsupervised Learning** Clustering Introduction-Mixture Densities, K-Means Clustering- Expectation-Maximization algorithm-Mixtures of Latent Varaible Models-Supervised Learning after Clustering-Spectral Clustering-Hierachial Clustering-Clustering- Choosing the number of Clusters.	CO2, CO3

	Lab Exercises 1. K means clustering 2. Hierarchical clustering **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** E. Alpaydin, Introduction to Machine Learning, 3rd Edition, MIT Press, 2014.	

Essential References

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

Recommended References

[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.

MDS332A: CATEGORICAL DATA ANALYSIS

Total Teaching Hours for Semester: 45

No of hours per week: 4L-0T-0P

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	LRNG Needs
CO1	Describe the categorical response	Global
CO2	Identify tests for contingency tables	**National**
CO3	Apply regression models for categorical response variables	Global
CO4	Analyse contingency tables using log-linear models	Global

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2		1		

CO2	2	2			1	2
CO3	3		2		1	
CO4	3	1				2

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	**CIA1 (20 MARKS)**	**CIA2 (50 MARKS)**	**CIA3 (20 MARKS)**	**ESE (100 MARKS)**
CO1	10	10		25
CO2	10	20		25
CO3		20	10	25
CO4			10	25

CO-UNIT MAPPING:

UNIT	**TOPICS/ SUB TOPICS**	**CO'S MAPPED**
UNIT 1 **Teaching Hours:9**	**Introduction** Categorical response data - Probability distributions for categorical data - Statistical inference for discrete data **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Agresti, A. (2012). *Categorical Data Analysis*, 3rd edition. New York, Wiley	CO1
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	CO2

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Agresti, A. (2012). *Categorical Data Analysis*, 3rd edition. New York, Wiley	
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 **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Agresti, A. (2012). *Categorical Data Analysis*, 3rd edition. New York, Wiley .	CO1,CO3
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 **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Agresti, A. (2012). *Categorical Data Analysis*, 3rd edition. New York, Wiley	CO3
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.	CO2,CO3,CO4

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Agresti, A. (2012). *Categorical Data Analysis*, 3rd edition. New York, Wiley.	

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.

MDS332B: MULTIVARIATE ANALYSIS

Total Teaching Hours for Semester: 45

No of hours per week: 4L-0T-0P

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	LRNG Needs
CO1	Understand multivariate data structure, multinomial, and multivariate normal distribution.	Global
CO2	Apply likelihood Ratio tests for multivariate normal proportions	National
CO3	Analyze multivariate data using (MANOVA) of one and two-way classified data.	National

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**					

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2				

CO2	2	2		1	2	2
CO3	2	2			1	2

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	**CIA1 (20 MARKS)**	**CIA2 (50 MARKS)**	**CIA3 (20 MARKS)**	**ESE (100 MARKS)**
CO1	10	15		30
CO2	10	20	10	30
CO3		15	10	40

CO-UNIT MAPPING:

UNIT	**TOPICS/ SUB TOPICS**	**CO'S MAPPED**
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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Anderson, T.W. 2009. An Introduction to Multivariate Statistical Analysis, 3rd Edition, John Wiley.	CO1
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.	CO1

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Anderson, T.W. 2009. An Introduction to Multivariate Statistical Analysis, 3rd Edition, John Wiley.	
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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Everitt B, Hothorn T, 2011. An Introduction to Applied Multivariate Analysis with R, Springer.	CO1, CO2,CO3
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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Everitt B, Hothorn T, 2011. An Introduction to Applied Multivariate Analysis with R, Springer.	CO1,CO2, CO3
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.	CO1,CO2, CO3

	Teaching /learning Strategy: Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** Barry J. Babin, Hair, Rolph E Anderson, and William C. Blac, 2013, Multivariate Data Analysis, Pearson New International Edition.	

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.

MDS332C: 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	LRNG Needs
CO1	Understand and apply the types of stochastic processes in various real-life scenarios.	**Global**
CO2	Demonstrate a discrete space stochastic process in a discrete index and estimate the evolving time in a state.	**Global**
CO3	Apply probability arguments to model and estimate the counts in continuous time.	**Global**
CO4	Evaluate the extinction probabilities of a generation.	**Global**
CO5	Development of renewal equations in discrete and continuous time.	**Global**
CO6	Understand the stationary process and application in Time Series Modelling	**Global**

Cross Cutting Issues:

Employability	Skill Development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**				**Yes**	

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6
CO1	1	2	2	2	3	-
CO2	3	3	2	3	1	3
CO3	3	3	2	3	3	3
CO4	3	3	2	3	1	3
CO5	3	3	2	3	2	-
CO6	3	3	2	3	3	3

CO-ASSESSMENT MAPPING-THEORY COMPONENT:

Course Outcomes /Unit	CIA I (20 MARKS)	CIA II (50 MARKS)	CIA III (20 MARKS)	ESE (100 MARKS)
CO1	5	5		10
CO2	15	10		10
CO3		15		20
CO4		20		20
CO5			10	20
CO6			5	20

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	*CO'S MAPPED
UNIT 1 **Teaching Hours: 09L**	**INTRODUCTION TO STOCHASTIC PROCESSES**	**CO1, CO2**

	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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Stochastic Processes, R.G Gallager, Cambridge University Press, 2013. 2. Stochastic Processes, S.M Ross, Wiley India Pvt. Ltd, 2008.	
UNIT 2 **Teaching Hours:09L**	**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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Stochastic Processes, R.G Gallager, Cambridge University Press, 2013. 2. Stochastic Processes, S.M Ross, Wiley India Pvt. Ltd, 2008.	**CO3**
UNIT 3 **Teaching Hours: 09L**	**BRANCHING PROCESS** Branching Processes – Galton – Watson Branching Process - Properties of Generating Functions – Extinction Probabilities – Distribution of Total Number of Progeny. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Stochastic Processes, R.G Gallager, Cambridge University Press, 2013.	**CO4**

	2. Stochastic Processes, S.M Ross, Wiley India Pvt. Ltd, 2008.	
UNIT 4 **Teaching Hours: 09L**	**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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Stochastic Processes, R.G Gallager, Cambridge University Press, 2013. 2. Stochastic Processes, S.M Ross, Wiley India Pvt. Ltd, 2008.	**CO5**
UNIT 5 **Teaching Hours: 09L**	**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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading** 1. Stochastic Processes, R.G Gallager, Cambridge University Press, 2013. 2. Stochastic Processes, S.M Ross, Wiley India Pvt. Ltd, 2008.	**CO6**

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.

MDSVAC1: CLOUD ESSENTIALS

Total Teaching Hours for Semester: 30

No of hours per week: 2L-0T-0P

Max Marks: 50 **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	LRNG Needs
CO1	Understand the *core concepts* of the cloud computing paradigm.	Global
CO2	Apply fundamental concepts of cloud *infrastructures, cloud storage* and in storage systems such as Amazon S3 and HDFS.	Global
CO3	Analyze various *cloud programming models* and apply them to solve problems on the cloud.	Global

Cross Cutting Issues:

Employability	Skill development	Entrepreneurship	Gender	Environment	Sustainability	Human Values and Professional Ethics
Yes	**Yes**				**Yes**	**Yes**

CO-PO MAPPING:

Course Outcomes /Programme Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	3	1	1	1	1	1	2
CO2	3	2	3	2	2	3	2
CO3	3	3	3	3	3	3	2

CO-ASSESSMENT MAPPING:

Course Outcomes /Unit	CIA1 (25 MARKS)	CIA2 (25 MARKS)	ESE (50 MARKS)
CO1	10	5	10
CO2	15	10	20
CO3		10	20

CO-UNIT MAPPING:

UNIT	TOPICS/ SUB TOPICS	CO'S MAPPED
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 **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **Essential Reading:** [1].Douglas Corner The Cloud Computing Book: The Future of Computing Explained,CRC Press,2021	CO1

	[2].Chellammal Surianarayanan,Essentials of Cloud Computing: A Holistic Perspective,Springer,2019	
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. **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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	CO1
UNIT 3 **Teaching Hours:6**	**Virtualization:** Virtualization (CPU, Memory, I/O),Case Study: Amazon EC2,Software Defined Networks (SDN),Software Defined Storage (SDS) **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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	CO1, CO2
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,	CO1, CO2

	DynamoDB) ,Cloud Object Storage (Amazon S3, OpenStack Swift, Ceph) **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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	
UNIT 5 **Teaching Hours:6**	**Programming Models:** Distributed Programming for the Cloud Data-Parallel Analytics with Hadoop MapReduce (YARN) **Teaching /learning Strategy:** Lecture /Discussion/Presentation/Problem solving/Class Activity **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	CO1, CO2, CO3

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

2022 BATCH MSc Data Science
PROGRAMME STRUCTURE

1 Semester

Course Code	Course Title	Hours Per Week	Credits	Marks
MDS131	MATHEMATICAL FOUNDATION FOR DATA SCIENCE - I	4	4	100
MDS132	PROBABILITY AND DISTRIBUTION THEORY	4	4	100
MDS133	PRINCIPLES OF DATA SCIENCE	4	4	100
MDS134	RESEARCH METHODOLOGY	2	2	50
Choose Any One (Foundational Elective)				
MDS161A	INTRODUCTION TO STATISTICS	2	2	50
MDS161B	INTRODUCTION TO COMPUTERS AND PROGRAMMING			
MDS161C	LINUX ADMINISTRATION			
MDS171	DATA BASE TECHNOLOGIES	6	5	150
MDS172	INFERENTIAL STATISTICS	6	5	150
MDS173	PROGRAMMING FOR DATA SCIENCE IN PYTHON	6	4	100
HOLODD	HOLISTIC EDUCATION	1	-	-
Total	**-**	**35**	**30**	**800**

2 Semester

Course Code	Course Title	Hours Per Week	Credits	Marks
MDS231	MATHEMATICAL FOUNDATION FOR DATA SCIENCE - II	4	4	100
MDS232	REGRESSION ANALYSIS	4	4	100
Choose Any One (Statistics Elective)				
MDS241A	MULTIVARIATE ANALYSIS	4	4	100

MDS241B	STOCHASTIC PROCESS			
MDS241C	CATEGORICAL DATA ANALYSIS			
MDS271	MACHINE LEARNING	6	5	150
Choose Any One (Computer Science Elective)				
MDS272A	WEB ANALYTICS	6	5	150
MDS272B	IMAGE AND VIDEO ANALYTICS			
MDS272C	INTERNET OF THINGS			
MDS273	JAVA PROGRAMMING	5	4	100
MDS281	SEMINAR	2	1	50
HOLEVEN	HOLISTIC EDUCATION	1	2	-
Total	**-**	**32**	**29**	**750**

3 Semester

Course Code	Course Title	Hours Per Week	Credits	Marks
MDS331	NEURAL NETWORKS AND DEEP LEARNING	4	4	100
Choose Any One (Statistics Elective)				
MDS341A	TIME SERIES ANALYSIS AND FORECASTING TECHNIQUES	4	4	100
MDS341B	BAYESIAN INFERENCE			
MDS341C	ECONOMETRICS			
MDS341D	BIO-STATISTICS			
MDS371	CLOUD ANALYTICS	6	5	150
MDS372	BUSINESS INTELLIGENCE	5	4	100
Choose Any One (Computer Science Elective)				
MDS373A	NATURAL LANGUAGE PROCESSING	6	5	150

MDS373B	HADOOP			
MDS373C	BIO INFORMATICS			
MDS373D	EVOLUTIONARY ALGORITHMS			
MDS373E	OPTIMIZATION TECHNIQUES			
MDS381	SPECIALIZATION PROJECT	4	2	100
VDSS311	PROGRAMMING FOR DATA SCIENCE IN R	2	-	
Total	**-**	**31**	**24**	**700**

4 Semester

Course Code	Course Title	Hours Per Week	Credits	Marks
MDS481	INDUSTRY PROJECT	2	12	300
Total	**-**	**2**	**12**	**300**

MDS331-NEURAL NETWORKS AND DEEP LEARNING

Total Teaching Hours For Semester:60 No of Lecture Hours/Week:4
Max Marks:100 Credits:4

Course Description and Course Objectives

The main aim of this course is to provide fundamental knowledge of neural networks and deep learning. 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

CO1: Understand the major technology trends in neural networks and deep learning
CO2: Build, train and apply neural networks and fully connected deep neural networks CO3: Implement efficient (vectorized) neural networks for real time application

Unit-1Teaching Hours:12

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.

Unit-2 Teaching Hours:12

SUPERVISED LEARNING NETWORK
Shallow neural networks- Perceptron Networks-Theory-Perceptron Learning RuleArchitecture-Flowchart for training Process-Perceptron Training Algorithm for Single and Multiple Output Classes.
Back Propagation Network- Theory-Architecture-Flowchart for training process-Training Algorithm-Learning Factors for Back-Propagation Network.
Radial Basis Function Network RBFN: Theory, Architecture, Flowchart and Algorithm.

Unit-3 Teaching Hours:12

CONVOLUTIONAL NEURAL NETWORK
Introduction - Components of CNN Architecture - Rectified Linear Unit (ReLU) Layer - Exponential Linear Unit (ELU, or SELU) - Unique Properties of CNN -Architectures of CNN -Applications of CNN.

Unit-4 Teaching Hours:12

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

Unit-5 Teaching Hours:12

AUTO ENCODER AND RESTRICTED BOLTZMANN MACHINE Introduction - Features of Auto encoder Types of Autoencoder Restricted Boltzmann Machine- Boltzmann Machine - RBM Architecture -Example - Types of RBM.

Essential References

1. S.N.Sivanandam, S. N. Deepa, Principles of Soft Computing, Wiley-India, 3rd Edition, 2018.
2. Dr. S Lovelyn Rose, Dr. L Ashok Kumar, Dr. D Karthika Renuka, Deep Learning Using Python, Wiley-India, 1st Edition, 2019.

Recommended References

1. Charu C. Aggarwal, Neural Networks and Deep Learning, Springer, September 2018. 2. Francois Chollet, Deep Learning with Python, Manning Publications; 1st edition, 2017
3. John D. Kelleher, Deep Learning (MIT Press Essential Knowledge series), The MIT Press, 2019.

Evaluation Pattern

CIA: 50%
ESE: 50%

MDS341A-TIME SERIES ANALYSIS AND FORECASTING TECHNIQUES
Total Teaching Hours For Semester:60 No of Lecture Hours/Week:4 Max Marks:100 Credits:4

Course Description and Course Objectives

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

CO1: Ability to approach and analyze univariate time series
CO2: Able 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 AND STOCHASTIC PROCESS 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, Holt- Winter smoothing, Forecasting based on smoothing.

Unit-2 Teaching Hours:12

STATIONARY TIME SERIES MODELS
Wold representation of linear stationary processes, 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, 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

STATE SPACE MODELS
Filtering, smoothing and forecasting using state space models, Kalman smoother, Maximum likelihood estimation, Missing data modifications

Essential References

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

Recommended References

1. Anderson T.W,Statistical Analysis of Time Series, John Wiley& Sons, Inc., New Jersey, 1971.
2. Shumway R.H and Stoffer D.S, Time Series Analysis and its Applications with R Examples, Springer, 2011.
3. P. J. Brockwell and R. A. Davis, Times series: Theory and Methods, 2nd Edition, Springer-Verlag, 2009.
4. S.C. Gupta and V.K. Kapoor, Fundamentals of Applied Statistics, 4th Edition, Sultan Chand and Sons, 2008.

Additional Information

NA

Evaluation Pattern

CIA: 50%
ESE: 50%

MDS341B-BAYESIAN INFERENCE

Total Teaching Hours For Semester:60 No of Lecture Hours/Week:4
Max Marks:100 Credits:4

Course Description and Course Objectives

To equip the students with the knowledge of conceptual, computational, and practical methods of Bayesian data analysis.

Course Outcomes

CO1: Understand Bayesian models and their specific model assumptions.
CO2: Identify suitable informative and non-informative prior distributions to derive posterior distributions.
CO3: Apply computer intensive methods like MCMC for approximating the posterior distribution.
CO4: Analyse the results obtained by Bayesian methods.

Unit-1Teaching Hours:12

INTRODUCTION
Basics on minimaxity: subjective and frequents probability, Bayesian inference, Bayesian estimation , prior distributions, posterior distribution, loss function, principle of minimum expected posterior loss, quadratic and other common loss functions, Advantages of being a Bayesian HPD confidence intervals, testing, credible intervals, prediction of a future observation.

Unit-2 Teaching Hours:12

BAYESIAN ANALYSIS WITH PRIOR INFORMATION
Robustness and sensitivity, classes of priors, conjugate class, neighbourhood class, density ratio class different methods of objective priors: Jeffrey's prior, probability matching prior, conjugate priors and mixtures, posterior robustness: measures and techniques

Unit-3 Teaching Hours:12

MULTIPARAMETER AND MULTIVARIABLE MODELS
Basics of decision theory, multi-parameter models, Multivariate models, linear regression, asymptotic approximation to posterior distributions

Unit-4 Teaching Hours:12

MODEL SELECTION AND HYPOTHESIS TESTING
Selection criteria and testing of hypothesis based on objective probabilities and Bayes' factors, large sample methods: limit of posterior distribution, consistency of posterior distribution, asymptotic normality of posterior distribution.

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, convergence issues

Essential References

1. Albert Jim (2009) Bayesian Computation with R, second edition, Springer, New York 2. Bolstad W. M. and Curran, J.M. (2016) Introduction to Bayesian Statistics 3rd Ed. Wiley, New York
3. Christensen R. Johnson, W. Branscum A. and Hanson T.E. (2011) Bayesian Ideas and data analysis : A introduction for scientist and Statisticians, Chapman and Hall, LondonA. Gelman, J.B. Carlin, H.S. Stern and D.B. Rubin (2004). Bayesian Data Analysis,2nd Ed. Chapman & Hall

Recommended References

1. Congdon P. (2006) Bayesian Statistical Modeling, Wiley, New York. 2. Ghosh, J.K. Delampady M. and T. Samantha (2006). An Introduction to Bayesian Analysis: Theory and Methods, Springer, New York.
3. Lee P.M. (2012) Bayesian Statistics: An Introduction-4th Ed. Hodder Arnold, New York.
4. Rao C.R. Day D. (2006) Bayesian Thinking, Modeling and Computation, Handbook of Statistics, Vol.25.

Additional Information

NA

Evaluation Pattern

CIA: 50%
ESE: 50%

MDS341C-ECONOMETRICS

Total Teaching Hours For Semester:60 No of Lecture Hours/Week:4
Max Marks:100 Credits:4

Course Description and Course Objectives

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 Outcomes

CO1: Demonstrate Simple and multiple Econometric models
CO2: Interpret the models adequacy through various methods
CO3: Demonstrate simultaneous Linear Equations model.

Unit-1 Teaching Hours:15

INTRODUCTION
Introduction to Econometrics- Meaning and Scope – Methodology of Econometrics – Nature and Sources of Data for Econometric analysis – Types of Econometrics

Unit-2 Teaching Hours:15

CORRELATION
Aitken's Generalised Least Squares(GLS) Estimator, Heteroscedasticity, Auto-correlation, Multicollinearity, Auto-Correlation, Test of Auto-correlation, Multicollinearity, Tools for Handling Multicollinearity

Unit-3 Teaching Hours:15

REGRESSION
Linear Regression with Stochastic Regressors, Errors in Variable Models and Instrumental Variable Estimation, Independent Stochastic linear Regression, Auto regression, Linear regression, Lag Models

Unit-4 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

Essential References

1. Johnston, J. (1997). Econometric Methods, Fourth Edition, McGraw Hill 2. Gujarathi, D., and Porter, D. (2008). Basic Econometrics, Fifth Edition, McGraw Hill

Recommended References

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.

Additional Information

NA

Evaluation Pattern

CIA : 50%
ESE : 50%

MDS341D-BIO-STATISTICS

Total Teaching Hours For Semester:60 No of Lecture Hours/Week:4
Max Marks:100 Credits:4

Course Description and Course Objectives

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

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: Interpret the measures of central tendency and measures of dispersion. CO4: Interpret the data based on the discrete and continuous probability distributions.
CO5: Apply parametric and non-parametric methods of statistical data analysis.

Unit-1 Teaching Hours:12

INTRODUCTION TO BIOSTATISTICS
Presentation of data - graphical and numerical representations of data - Types of variables, measures of location - dispersion and correlation - inferential statistics - probability and distributions - Binomial, Poisson, Negative Binomial, Hyper geometric and normal distribution.

Unit-2 Teaching Hours:12

PARAMETRIC AND NON - PARAMETRIC METHODS
Parametric methods - 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 - Pearson correlation coefficient - Non parametric methods: Chi-square test of independence and goodness of fit - Mann Whitney U test - Wilcoxon signed-rank test - Kruskal Wallis test - Friedman's test - Spearman's correlation test.

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

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 References

1. Marcello Pagano and Kimberlee Gauvreau (2018), Principles of Biostatistics, 2nd Edition, Chapman and Hall/CRC press
2. David Moore S. and George McCabe P., (2017) Introduction to practice of statistics, 9th Edition, W. H. Freeman.
3. Sundar Rao and Richard J., (2012) Introduction to Biostatistics and research methods, PHI Learning Private limited, New Delhi

Recommended References

1. Abhaya Indrayan and Rajeev Kumar M., (2018) Medical Biostatistics, 4th Edition, Chapman and Hall/CRC Press.
2. Gordis Leon (2018), Epidemiology, 6th Edition, Elsevier, Philadelphia 3. Ram, F. and Pathak K. B., (2016): Techniques of Demographic Analysis, Himalaya Publishing house, Bombay.
4. Park K., (2019), Park's Text Book of Preventive and Social Medicine, Banarsidas Bhanot, Jabalpur.

Additional Information

-

Evaluation Pattern

CIA:50%
ESE:50%

MDS371-CLOUD ANALYTICS

Total Teaching Hours For Semester:90 No of Lecture Hours/Week:6
Max Marks:150 Credits:5

Course Description and Course Objectives

The objective of this course is to explore the basics of cloud analytics and the major cloud solutions. Students will learn how to analyze extremely large data sets, and to create visual representations of that data. Also aim to provide students with hands-on experience working with data at scale.

Course Outcomes

CO1: Interpret the deployment and service models of cloud applications. CO2: Describe big data analytical concepts.
CO3: Ingest, store, and secure data.
CO4: Process and Visualize structured and unstructured data.

Unit-1Teaching Hours:18

INTRODUCTION
Introduction to cloud computing - Major benefits of cloud computing - Cloud computing deployment models - Private cloud - Public cloud - Hybrid cloud - Types of cloud computing services -Infrastructure as a Service – PaaS – SaaS - Emerging cloud technologies and services - Different ways to secure the cloud - Risks and challenges with the cloud - What is cloud analytics? Parameters before adopting cloud strategy - Technologies utilized by cloud computing

1. Creating Virtual Machines using Hypervisors
2. IaaS: Compute service - Creating and running Virtual Machines

Unit-2 Teaching Hours:18

CLOUD ENABLING TECHNOLOGIES
Virtualization - Load Balancing - Scalability & Elasticity – Deployment –Replication – Monitoring - Software Defined Networking - Network Function Virtualization – MapReduce - Identity and Access Management - Service Level Agreements - Billing 1. Storage as a Service: Ingesting & Querying data into cloud
2. Database as a Service: Building DB Server

Unit-3 Teaching Hours:18

BASIC CLOUD SERVICES & PLATFORMS
Compute Services
Amazon Elastic Compute Cloud - Google Compute Engine
- Windows Azure Virtual Machines
Storage Services
Amazon Simple Storage Service - Google Cloud Storage - Windows Azure Storage
Database Services
Amazon Relational Data Store - Amazon DynamoDB - Google Cloud SQL - Google Cloud Datastore - Windows Azure SQL Database - Windows Azure Table Service 1. PaaS: Working with GoogleAppEngine

Unit-4 Teaching Hours:18

DATA INGESTION AND STORING
Cloud Dataflow - The Dataflow programming model - Cloud Pub/Sub - Cloud storage - Cloud SQL - Cloud BigTable - Cloud Spanner - Cloud Datastore - Persistent disks 1. Database as a Service: Building DB Server
2. Transforming data
PROCESSING AND VISUALIZING
Google BigQuery - Cloud Dataproc - Google Cloud Datalab - Google Data Studio 1. Visualize structured data and unstructureddata

Unit-5 Teaching Hours:18

MACHINE LEARNING, DEEP LEARNING AND AI
Services on Artificial intelligence - Machine learning - Cloud Natural Language API – TensorFlow - Cloud Speech API - Cloud Translation API - Cloud Vision API - Cloud Video Intelligence – Dialogflow – AutoML
1. Load and query data in a data warehouse
2. Setting up and executing a data pipeline job to load data into cloud

Essential References

1. Sanket Thodge, Cloud Analytics with Google Cloud Platform, Packt Publishing, 20 18.
2. Arshdeep Bahga and Vijay Madisetti, Cloud computing - A Hands-On Approach, Create Space Independent Publishing Platform, 2014.

Recommended References

1. Deven Shah, Kailash Jayaswal, Donald J. Houde, Jagannath Kallakurchi, Cloud Computing - Black Book, Wiley, 2014.
2. Thomas Erl, Ricardo Puttini, Zaigham Mahmood, Cloud Computing: Concepts, Technology & Architecture, Prentice Hall, 2014.

Additional Information

NA

MDS372 BUSINESS INTELLIGENCE

Total Teaching Hours for Semester: 75
Max Marks: 100 Credits: 04

Course Objectives

This course is designed to introduce students the concepts of business intelligence and also provide students with an understanding of data warehousing and data mining along with associated tools and techniques and their benefits to organizations of all sizes.

Course Outcomes

CO1: Understand the fundamentals of Business Intelligence and Analytics

CO2:Apply theconcepts of data warehouse concepts required for Business Intelligence

CO3:Build a performance dashboard using data visualization and visual analytics.

CO4:Implement the business intelligence perspective of data mining and text mining concepts.

Unit – 1 Teaching Hours: 15

An Overview of Business Intelligence, Analytics, andDecision Support: Changing Business Environments and Computerized Decision Support - A Framework for Business Intelligence (BI) - Transaction Processing VERSUS Analytic Processing - Successful BI Implementation - Business Analytics Overview: Descriptive Analytics - Predictive Analytics - Prescriptive Analytics - BriefIntroduction to Big Data Analytics Applications of BI.

Lab Exercises:
1.Case study on Transaction Processing.
2.Case Study on Predictive Analytics

Unit – 2 Teaching Hours: 15 Business Intelligence Tools and Applications:

Ad hoc Analysis-Online Analytical Processing-Mobile BI-Real-time BI-Operation Intelligence
Open-Source BI-Embedded BI-Collaborative BI-Location Intelligence-Business intelligence
vendors and market

Lab Exercises:
1.Exercise on OLAP in BI.
2. Exercise on Real time BI.

Unit – 3 Teaching Hours: 15 Power BI

Power BI Overview-Installation-Data Sources-Query Editor-Importing Files-Data Modeling Lookup Data Tables-Active vs. Inactive Relationships-Roles-Refreshing Data and Hierarchies-Data Modeling-DAX-Calculated Columns-Measures-Design and Interactive Reports-Dashboard

Lab Exercises:
1.Exercise on Data modeling in Power BI
2. Exercise on Dashboards & Reports in Power BI.

Unit – 4 Teaching Hours: 15

Tableau Basics

Tableau Overview-Data Sources, First Bar Chart Graph-the Extracted Data- Knowledge of Aggregation, Granularity, and Time-Series-Working with Charts and Filter-Overview of First Dashboard, Maps, and Scatter Plots-Joins and Relationship- Data Joining- Map Creation.

Lab Exercises:
1.Exercise on Extracted Data in Tableau.
2.Exercise on Joins, Maps, Plots in Tableau.

Unit – 5 Teaching Hours: 15
Working with Tableau

First Dashboard Creation with Highlighting and Filters-Overview of Dual-axis Chart, Joining, Relationship, and Blending-Joining with Different Conditions, i.e., Multiple Fields and Duplicate Values-Working on Blending Data-Creation of Dual Axis Chart Understanding of Calculated Fields-Understanding of Relationship Data-Overview of New Dashboard-Updated Way of Data Preparation-Overview of New Design Feature and Many More-Advancement in Tableau

Lab Exercises:
1.Exercise on Dashboard Creation and Blending in Tableau.
2. Exercise on Data preparation, Data relationship and fields in Tableau.

Essential Readings:

1. Chandraish Sinha (2022).” Mastering Power BI”,1st Edition, BPB Publications.
2. Marleen, David,” Mastering Tableau 2021: Implement advanced businessintelligence techniques and analytics with Tableau”, 3rd Edition, Pakt,
3. Ramesh Sharda, Dursun, Delen, Efraim Turban (2017). “Business Intelligence: Manegerial Perspective on Analytics”, 3rd Edition, Pearson Publication.

Recommended Readings:

1. Ahmed Sherif(2016).”Practical Business Intelligence”,Packt Publishing.

MDS373A: Natural Language Processing

Total Teaching Hours for Semester: 90
Max Marks: 150 Credits:05 Course Objectives

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

Course Outcomes

CO1: Understand various approaches on syntax and semantics in NLP

CO2: Apply various methods to discourse, generation, dialogue and summarization using NLP.

CO3: Analyze various methodologies used in Machine Translation, machine learning techniques used in NLP including unsupervised models and to analyze real time applications

Unit-1 Teaching Hours:12 INTRODUCTION

Introduction to NLP- Background and overview- NLP Applications -NLP hard Ambiguity Algorithms and models, Knowledge Bottlenecks in NLP- Introduction to NLTK, Case study

Unit-2 Teaching Hours:12 PARSING AND SYNTAX

Word Level Analysis: Regular Expressions, Text Normalization, Edit Distance, Parsing and Syntax- Spelling, Error Detection and correction-Words and Word classes- Part-of Speech Tagging, Naive Bayes and Sentiment Classification: Case study

Unit-3 Teaching Hours:12 SMOOTHED ESTIMATION AND LANGUAGE MODELLING

N-gram Language Models: N-Grams, Evaluating Language Models -The language modelling problem

SEMANTIC ANALYSIS AND DISCOURSE PROCESSING

Semantic Analysis: Meaning Representation-Lexical Semantics- Ambiguity-Word Sense Disambiguation. Discourse Processing: cohesion-Reference Resolution- Discourse Coherence and Structure.

Unit-4 Teaching Hours:12 NATURALLANGUAGE GENERATION AND MACHINE TRANSLATION Natural Language Generation: Architecture of NLG Systems, Applications

Machine Translation: Problems in Machine Translation- Machine Translation Approaches Evaluation of Machine Translation systems.

Case study: Characteristics of Indian Languages

Unit-5 Teaching Hours:12 INFORMATION RETRIEVAL AND LEXICAL RESOURCES

Information Retrieval: Design features of Information Retrieval Systems-Classical, Non classical, Alternative Models of Information Retrieval – valuation Lexical Resources: Word Embeddings - Word2vec- Glove.

UNSUPERVISED METHODS IN NLP Graphical Models for Sequence Labelling in NLP

Lab Exercises: Total Hours:30

1. Write a program to tokenize text
2. Write a program to count word frequency and to remove stop words
3. Write a program to program to tokenize Non-English Languages
4. Write a program to get synonyms from WordNet
5. Write a program to get Antonyms from WordNet
6. Write a program for stemming Non-English words
7. Write a program for lemmatizing words Using WordNet
8. Write a program to differentiate stemming and lemmatizing words
9. Write a program for POS Tagging or Word Embeddings.
10. Case study-based program (IBM) or Sentiment analysis

Essential Reading

1. Speech and Language Processing, Daniel Jurafsky and James H., 2nd Edition, Martin Prentice Hall,2013.
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, O'Reilly Media; 1 edition, 2009.

Web resources:

1. https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf
2. https://nptel.ac.in/courses/106101007/
3. NLTK – Natural Language Tool Kit- http://www.nltk.org

MDS373B-HADOOP

Total Teaching Hours For Semester:90 No of Lecture Hours/Week:6

Max Marks:150 Credits:5

Course Description and Course Objectives

The subject is intended to give the knowledge of Big Data evolving in every real-time applications 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) and HBase in Ubuntu platform.

Course Outcomes

CO1: Understand the Big Data concepts in real time scenario
CO2: Understand the big data systems and identify the main sources of Big Data in the real world.
CO3: Demonstrate an ability to use Hadoop framework for processing Big Data for Analytics.
CO4: Evaluate the Map reduce approach for different domain problems.

Unit-1 Teaching Hours:15

INTRODUCTION

Distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications, Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce.

Apache Hadoop– Moving Data in and out of Hadoop – Understanding inputs and outputs ofMapReduce - Data Serialization, Problems with traditional large-scale systems-Requirements for a new approach-Hadoop – Scaling-Distributed Framework-Hadoop v/s RDBMS-Brief history of Hadoop.

Lab Exercise

1. Installing and Configuring Hadoop

Unit-2 Teaching Hours:15

CONFIGURATIONS OF HADOOP

Hadoop Processes (NN, SNN, JT, DN, TT)-Temporary directory – UI-Common errors when running Hadoop cluster, solutions.Setting up Hadoop on a local Ubuntu host: Prerequisites, downloading Hadoop, setting up SSH, configuring the pseudo-distributed mode, HDFS directory, NameNode, Examples of MapReduce, Using Elastic MapReduce, Comparison of local versus EMR Hadoop.
Understanding MapReduce:Key/value pairs,TheHadoop Java API for MapReduce, Writing MapReduce programs, Hadoop-specific data types, Input/output.Developing MapReduce Programs: Using languages other than Java with Hadoop, Analysing a large dataset.

Lab Exercise

1. 1. Word count application in Hadoop.

2. 2. Sorting the data using MapReduce.

3. 3. Finding max and min value in Hadoop.

Unit-3 Teaching Hours:15

ADVANCED MAPREDUCE TECHNIQUES

Simple, advanced, and in-between Joins, Graph algorithms, using language-independent data structures.Hadoop configuration properties - Setting up a cluster, Cluster access control, managing the NameNode, Managing HDFS, MapReduce management, Scaling.

Lab Exercise:

1. Implementation of decision tree algorithms using MapReduce.

2. Implementation of K-means Clustering using MapReduce.

3. Generation of Frequent Itemset using MapReduce.

Unit-4 Teaching Hours:15

HADOOP STREAMINGHadoop Streaming -

Streaming Command Options -

Specifying a Java Class as the Mapper/Reducer - Packaging Files With Job Submissions - Specifying Other Plug-ins for Jobs.

Lab Exercise:

1. Count the number of missing and invalid values through joining two large given datasets.
2. Using hadoop's map-reduce, Evaluating Number of Products Sold in Each Country in the online shopping portal. Dataset is given.
3. Analyze the sentiment for product reviews, this work proposes a MapReduce technique provided by Apache Hadoop.

Unit-5 Teaching Hours:15 HIVE & PIG

Architecture, Installation, Configuration, Hive vs RDBMS, Tables, DDL & DML, Partitioning & Bucketing, Hive Web Interface, Pig, Use case of Pig, Pig Components, Data Model, Pig Latin.

Lab Exercise

1. Trend Analysis based on Access Pattern over Web Logs using Hadoop. 2. Service

Rating Prediction by Exploring Social Mobile Users Geographical Locations.

Unit-6 Teaching Hours:15 Hbase

RDBMS VsNoSQL, HBasics, Installation, Building an online query application – Schema design, Loading Data, Online Queries, Successful service.

Hands On: Single Node Hadoop Cluster Set up in any cloud service provider- How to create instance.How to connect that Instance Using putty.Installing Hadoop framework on this instance. Run sample programs which come with Hadoop framework.

Lab Exercise:

1. 1. Big Data Analytics Framework Based Simulated Performance and Operational Efficiencies Through Billons of Patient Records in Hospital System.

Essential References

1] Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Professional Hadoop Solutions, Wiley, 2015.

[2] Tom White, Hadoop: The Definitive Guide, O'Reilly Media Inc., 2015.

[3] Garry Turkington, Hadoop Beginner's Guide, Packt Publishing, 2013.

MDS373C-BIO INFORMATICS

Total Teaching Hours For Semester:90 No of Lecture Hours/Week:6
Max Marks:150 Credits:5

Course Description and Course Objectives

To enable the students to learn the information search and retrieval, Genome analysis and Gene mapping, alignment of multiple sequences, and PERL for Bioinformatics.

Course Outcomes

CO1: To understand the molecular Biology and Bioinformatics applications. CO2: Apply the modeling and simulation technologies in Biology and medicine. CO3: Evaluate the algorithms to find the similarity between protein and DNA sequences.

Unit-1 Teaching Hours:18

BIOINFORMATICS

Introduction, Historical Overview and Definition, Applications, Major databases in Bioinformatics, Data management and Analysis, Central Dogma of Molecular Biology.

INFORMATION SEARCH AND RETRIEVAL

Introduction, Tools for web search, Data retrieval tools, Data mining of Biological databases.

Lab Exercise

1. Test and verify the basic Linux commands and Filters.
2. Create the file(s) and verify the file handling commands.

Unit-2 Teaching Hours:18

GENOME ANALYSIS AND GENE MAPPING

GENOME ANALYSIS AND GENE MAPPING Introduction, Genome analysis, Genome mapping, Sequence assembly problem, Genetic mapping and linkage analysis, Physical maps, Cloning the entire Genome, Genome sequencing, Applications of Genetic maps, Identification of Genes in Contigs, Human Genome Project. ALIGNMENT OF PAIRS OF SEQUENCES Introduction, Biological motivation of alignment, Methods of sequence alignments, Using score matrices, Measuring sequence detection

Lab Exercise

1. Create directories and verify the directory commands.
2. Perform basic mathematical operations using PERL.
3. Write a PERL script to demonstrate the Array operations and Regular

expressions.

Unit-3 Teaching Hours:18

ALIGNMENT OF MULTIPLE SEQUENCES

ALIGNMENT OF MULTIPLE SEQUENCES Methods of multiple sequence alignment, Evaluating multiple alignments, Applications of multiple alignments, Phylogenetic analysis, Methods of phylogenetic analysis, Tree evaluation, Problems in Phylogenetic analysis.

TOOLS FOR SIMILARITY SEARCH AND SEQUENCE ALIGNMENT Introduction, Working with FASTA, Working with BLAST, Filtering and Gapped BLAST, FASTA and BLAST algorithm comparison.

Lab Exercise

1. Write a PERL script to concatenate DNA sequences.
2. Write a PERL script to transcribe DNA sequence into RNA sequence 3.

Write a PERL script to calculate the reverse complement of a strand of DNA.

Unit-4 Teaching Hours:18

PERL FOR BIOINFORMATICS

Sequences and Strings: Representing sequence data, Program to store a DNA sequence, Concatenating DNA fragments, Transcription DNA to RNA, Proteins, Files and Arrays, Reading Proteins in Files, Arrays, Scalar and List Context.

Motifs and Loops: Flow control, Code layout, Finding motifs, Counting Nucleotides, Exploding strings and arrays, Operating on strings. Subroutine and Bugs: Subroutines,Scoping and Subroutines, Command line * arguments and Arrays, Passing data toSubroutines, Modules and Libraries of Subroutines.

Lab Exercise

1. Write a PERL script to read protein sequence data from a file.
2. Write a PERL script to search for a motif in a DNA sequence.

Unit-5 Teaching Hours:18

THE GENETIC CODE

Hashes, Data structure and algorithms for Biology, Translating DNA into Proteins, Reading DNA from the files in FASTA format, Reading Frames. GenBank: GenBank files, GenBank Libraries, Separating Sequence and Annotation, Parsing Annotations, Indexing GenBank with DBM. Protein Data Bank: Files and Folders, PDB Files, Parsing PDB Files.

1. Write a PERL script to append ACGT to DNA using a subroutine.

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

Essential References

[1] Bioinformatics: Methods and Applications, S. C. Rastogi, Namita Mendirata and Parag Rastogi, 4th Edition, PHI Learning, 2013.

[2] Beginning Perl for Bioinformatics, Tisdall James, 1st edition, Shroff Publishers (O'Reilly), 2009.

Recommended References

[1] Introduction to Bioinformatics, Arthur M Lesk, 2nd Edition, Oxford University Press,4th edition, 2014.

[2] Bioinformatics Technologies, Yi-Ping Phoebe Chen (Ed), 1st edition, Springer, 2005.

[3] Bioinformatics Computing, Bryan Bergeron, 2nd Edition, Prentice Hall, 1st edition, 2003.

Web resources:

[1] http://cac.annauniv.edu/PhpProject1/aidetails/afug_2013_fu/24.%20BIO%20MED.pdf

[2] https://www.amrita.edu/school/biotechnology/academics/pg/introduction bioinformaticsbif410

[3] https://canvas.harvard.edu/courses/8084/assignments/syllabus

[4] https://www.coursera.org/specializations/bioinformatics

[5]http://www.dtc.ox.ac.uk/modules/introduction-bioinformatics-bioscientists.html Evaluation Pattern

CIA 50%

ESE 50%

MDS373D-EVOLUTIONARY ALGORITHMS

Total Teaching Hours For Semester:90 No of Lecture Hours/Week:6
Max Marks:150 Credits:5

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

CO1:Basic understanding of evolutionary computing concepts and techniques
CO2:Classifyrelevantreal-time problems for the applications of evolutionary algorithms CO3:Design solutions using evolutionary algorithms

Unit-1Teaching Hours:18

INTRODUCTION TO EVOLUTIONARY COMPTUTING
Terminologies – Notations – Problems to be solved – Optimization – Modeling – Simulation
– Search problems – Optimization constraints
Lab Program
1. Implementation of single and multi-objectivefunctions
2. Implementation of binaryGA

Unit-2 Teaching Hours:18

EVOLUTIONARY PROGRAMMING
Continuous evolutionary programming – Finite state machine optimization – Discrete evolutionary programming – The Prisoner's dilemma
EVOLUTION STRATEGY
One plus one evolution strategy – The 1/5 Rule – (μ+1) evolution strategy – Self adaptive evolution strategy
Lab Program
1. Implementation of continuousGA
2. Implementation of evolutionaryprogramming

Unit-3 Teaching Hours:18

GENETIC PROGRAMMING
Fundamentals of genetic programming – Genetic programming for minimal time control
EVOLUTIONARY ALGORITHM VARIATION
Initialization – Convergence – Population diversity – Selection option – Recombination – Mutation
Lab Program
1. Implementation of geneticprogramming
2. Implementation of Ant ColonyOptimization

Unit-4 Teaching Hours:18

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

Unit-5 Teaching Hours:18

MULT-OBJECTIVE OPTIMIATION
Pareto Optimality – Hyper volume – Relative coverage – Non-pareto based EAs – Pareto based EAs – Multi-objective Biogeography based optimization
Lab Program
1. Simulation of EA in Planning problems (routing, scheduling, packing) and Design problems (Circuit, structure,art)
2. Simulation of EA in classification/predictionmodelling

Essential References

[1] D. Simon, Evolutionary optimization algorithms: biologically inspired and population-based approaches to computer intelligence. New Jersey: John Wiley, 2013.

Recommended References

1. Eiben and J. Smith, Introduction to evolutionary computing. 2nd ed. Berlin: Springer, 2015.
2. D.Goldberg,Geneticalgorithmsinsearch,optimization,andmachinelearning.Boston: Addison-Wesley,2012.
3. K. Deb, Multi-objective optimization using evolutionary algorithms. Chichester: John Wiley & Sons,2009.
4. R. Poli, W. Langdon, N. McPhee and J. Koza, A field guide to genetic programming. [S.l.]: Lulu Press,2008.
5. T.Bäck,Evolutionaryalgorithmsintheoryandpractice.NewYork:OxfordUniv.Press, 1996.
Web Resources:

1 E.A.EandS.J.E,"IntroductiontoEvolutionaryComputing|Theon-line accompaniment to the book Introduction toEvolutionary Computing",Evolutionarycomputation.org,2015.[Online].Available: http://www.evolutionarycomputation.org/.
2 F.Lobo,"EvolutionaryComputation2018/2019",Fernandolobo.info,2018.[Online]. Available:http://www.fernandolobo.info/ec1819.
3 "EClabTools",Cs.gmu.edu,2008.[Online].Available: https://cs.gmu.edu/~eclab/tools.html.

4 "Kanpur Genetic Algorithms Laboratory", Iitk.ac.in, 2008. [Online]. Available: https://www.iitk.ac.in/kangal/codes.shtml.
5 "Course webpage Evolutionary Algorithms", Liacs.leidenuniv.nl, 2017. [Online]. Available:http://liacs.leidenuniv.nl/~csnaco/EA/misc/ga_demo.htm.

Evaluation Pattern

CIA: 50%
ESE : 50%

MDS373E-OPTIMIZATION TECHNIQUES

Total Teaching Hours For Semester:90 No of Lecture Hours/Week:6

Max Marks:150 Credits:5

Course Description and Course Objectives

This course will help the students to acquire and demonstrate the implementation of the necessary algorithms for solving advanced level Optimization techniques.

Course Outcomes

CO1: Apply the notions of linear programming in solving transportation problems CO2: Understand the theory of games for solving simple games
CO3: Use linear programming in the formulation of the shortest route problem.
CO4: Apply algorithmic approach in solving various types of network problems
CO5: Create applications using dynamic programming.

Unit-1Teaching Hours:18

INTRODUCTION

Operations Research Methods - Solving the OR model - Queuing and Simulation models – Art of modelling – phases of OR study.

MODELLING WITH LINEAR PROGRAMMING

Two variable LP model – Graphical LP solution – Applications. Simplex method and sensitivity analysis – Duality and post-optimal Analysis- Formulation of the dual problem.

Lab Exercise

1. Simplex Method
2. Dual Simplex Method

Unit-2 Teaching Hours:18

TRANSPORTATION MODEL

Determination of the Starting Solution – Iterative computations of the transportation algorithm. Assignment Model: The Hungarian Method – Simplex explanation of the Hungarian Method – The trans-shipment Model.

Lab Exercise

1. Balanced Transportation Problem
2. Unbalanced Transportation Problem
3. Assignment Problems

Unit-3 Teaching Hours:18

NETWORK MODELS

Minimal Spanning tree Algorithm – Linear Programming formulation of the shortest-route problem. Maximal Flow Model: Enumeration of cuts – Maximal Flow Diagram – Linear Programming Formulation of Maximal Flow Model.

CPM and PERT

Network Representation – Critical Path Computations – Construction of the time Schedule – Linear Programming formulation of CPM – PERT networks.

Lab Exercise:

1. Shortest path computations in a network
2. Maximum flow problem

Unit-4 Teaching Hours:18

GAME THEORY

Strategic Games and examples - Nash equilibrium and examples - Optimal Solution of two person zero sum games - Solution of Mixed strategy games - Mixed strategy Nash equilibrium - Dominated action with example.

GOAL PROGRAMMING

Formulation – Tax Planning Problem – Goal Programming algorithms – Weights method – Preemptive method.

Lab Exercise:

1. Critical path Computations
2. Game Programming

Unit-5 Teaching Hours:18

MARKOV CHAINS

Definition – Absolute and n-step Transition Probability – Classification of states.

DYNAMIC PROGRAMMING

Recursive nature of computation in Dynamic Programming – Forward and Backward Recursion – Knapsack / Fly Away / Cargo-Loading Model – Equipment Replacement Model.

Lab Exercise:

1. Goal Programming
2. Dynamic Programming

Essential References

1. Hamdy A Taha, Operations Research, 9th Edition, Pearson Education, 2012. 2. Garrido José M. Introduction to Computational Models with Python. CRC Press, 2016.

Recommended References

1. Rathindra P Sen, Operations Research – Algorithms and Applications, PHI

Learning Pvt. Limited, 2011
2. R. Ravindran, D. T. Philips and J. J. Solberg, Operations Research: Principles and Practice, 2nd ed., John Wiley & Sons, 2007.
3. F. S. Hillier and G. J. Lieberman, Introduction to operations research, 8th ed., McGraw-Hill Higher Education, 2004.
4. K.C. Rao and S. L. Mishra, Operations research, Alpha Science International, 2005.
5. Hart, William E. Pyomo: Optimization Modeling in Python. Springer, 2012. 6. Martin J. Osborne, An introduction to Game theory, Oxford University Press, 2008

Additional Information

NA

Evaluation Pattern

CIA: 50%
ESE: 50%

MDS381-SPECIALIZATION PROJECT

Total Teaching Hours For Semester:60 No of Lecture Hours/Week:4
Max Marks:100 Credits:2

Course Description and Course Objectives

The course is designed to provide a real-world project development and deployment environment for the students.

Course Outcomes

CO1: Identify the problem and relevant analytics for the selected domain. CO2: Apply appropriate design/development strategy and tools.

Unit-1 Teaching Hours:60

Specialization Project
Project will be based on the specialization domains which students are opted for during this semester.

Essential References

-

Recommended References

-

Evaluation Pattern

CIA: 50%
ESE: 50%

VDS311--PROGRAMMING FOR DATA SCIENCE IN R Total Teaching Hours For Semester:30
No of Lecture Hours/Week:2 **MaxMarks:**
Credits:

Course Description and Course Objectives

This lab is designed to introduce implementation of practical machine learning algorithms using R programming language. The lab will extensively use datasets from real life situations.

Course Outcomes

CO1: Demonstrate to use R in any OS (Windows / Mac / Linux).
CO2: Analyse the use of basic functions of R Package.
CO3: Demonstrate exploratory data analysis (EDA) for a given data set.
CO4: Create and edit visualizations with R
CO5: Implement and assess relevance and effectiveness of machine learning algorithms for a given dataset.

Unit-1 Teaching Hours:6

R INSTALLTION, SETUP AND LINEAR REGRESSION

Download and install R – R IDE environments – Why R – Getting started with R – Vectors and Data Frames – Loading Data Frames – Data analysis with summary statistics and scatter plots – Summary tables - Working with Script Files
Linear Regression – Introduction – Regression model for one variable regression – Selecting best model – Error measures SSE, SST, RMSE, R2 – Interpreting R2 – Multiple linear regression – Lasso and ridge regression – Correlation – Recitation – A minimum of 3 data sets for practice

Unit-2 Teaching Hours:6

LOGISTIC REGRESSION

Logistic Regression – The Logit – Confusion matrix – sensitivity, specificity – ROC curve – Threshold selection with ROC curve – Making predictions – Area under the ROC curve (AUC) - Recitation – A minimum of 3 data sets for practice

Unit-3 Teaching Hours:6

DECISION TREES
Approaches to missing data – Data imputation – Multiple imputation – Classification and Regression Tress (CART) – CART with Cross Validation – Predictions from CART – ROC curve for CART – Random Forests – Building many trees – Parameter selection – K-fold Cross Validation – Recitation – A minimum of 3 data sets for practice

Unit-4 Teaching Hours:6

TEXT ANALYTICS AND NLP

Using text as data – Text analytics – Natural language processing – Bag of words – Stemming – word clouds – Recitation – min 3 data sets for practice – Time series analysis – Clustering – k-mean clustering – Random forest with clustering – Understanding cluster patterns – Impact of clustering – Heatmaps – Recitation – min 3 data sets for practice

Unit-5 Teaching Hours:6

ENSEMBLE MODELLING

Support Vector Machines – Gradient Boosting – Naive Bayes - Bayesian GLM – GLMNET - Ensemble modeling – Experimenting with all of the above approaches (Units 1-5) with and without data imputation and assessing predictive accuracy – Recitation – min 3 data sets for practice PROJECT – A concluding project work carried out individually for a common data set

Essential References

[1]. Statistics : An Introduction Using R, Michael J. Crawley, WILEY, Second Edition, 2015.

Recommended References

[1].Hands-on programming with R, Garrett Grolemund, O'Reilley, 1st Edition, 2014 [2]. R for everyone, Jared Lander, Pearson, 1st Edition, 2014

Evaluation Pattern

CIA - 50%
ESE - 50%