PERIYAR UNIVERSITY

PeriyarPalkalai Nagar, Salem-636011

(NAAC A Grade-STATE UNIVERSITY – NIRF RANK 83 – ATAL RANK 4)

DEPARTMENT OF COMPUTER SCIENCE

M.Sc. DEGREE

DATA SCIENCE

[Choice Based Credit System (CBCS)]

OBE REGULATIONS AND SYLLABUS

(Effective from the academic year 2020-21 and thereafter)

M. Sc. DATA SCIENCE

OBE REGULATIONS AND SYLLABUS
(With effect from the academic year 2020-2021 onwards)

1. Preamble

The **M.Sc-Data Science** programme is aimed at preparing data science professionals for research, academics, industry, design, and development in analytical technologies. It aims to reduce the gap between academic and industry by providing industry based expertise in the field of data science.

2. General Graduate Attributes

GA 1: The fundamental knowledge

- This refers to a group of concepts associated with the discipline of computer science and how they are applied in user interfaces.
- Initially, students learn about basic concepts of algorithms, programming language, and user interface.
- To develop in-depth knowledge about the data analytics concept through course work.

GA 2: Design and conduct experiments

- To apply the principles of data science to real-world problems.

GA 3: Software and tools

- A proficient user of existing computing tools.
- A proficient developer of new computing tools.
- A competent analyzer, processor, and reporter of simple and complex data.

GA 4: Solve computer science problems

- Be equipped with a range of fundamental principles of Computer Science that will provide the basis for future learning and enable them to adapt to the constant rapid development of the field.
- Be able to apply mathematics, logic, and statistics to the design, development, and analysis of software systems.
- Be able to apply ethical practices in handling data of any kind.

GA 5: To apply algorithmic principles

- Identifies the key intellectual themes of the field as algorithmic thinking, information representation, and computer programs.

GA 6: To acquire the latest technical skills

- To enable the students to acquire the latest technical skills and build their carrier based on continuous learning and adaptability.

GA 7: Theoretical and practical knowledge

- Theoretical knowledge refers to learning things, concept, principle, information regarding anything from books. It means that the theoretical knowledge is knowledge which is basic knowledge or concept of things.
- Practical knowledge and Practical skills are two different things. Practical knowledge is the knowledge in which we can apply and implement ideas using practices.
- Be able to apply algorithms such as clustering and classification to sample dataset
- To build machine intelligence algorithm for solving complex decision-making problems

GA 8: Leadership, initiative, and teamwork:

- Ability to work effectively in a team and lead in a multidisciplinary environment

GA 9: Creative:

- Demonstrates critical thinking, imagination, and intellectual agility.
- Strives to be innovative and experimental in advancing knowledge and in creating solutions.

GA 10: Specialist:

- Possesses breadth and depth of knowledge in their specialist area.
- Applies through research and inquiry a systematic and critical approach
- Analytical approach to identifying and resolving problems.

GA 11: Intellectual Rigour

- An ability to think clearly and deeply with rigour when faced with new knowledge and arguments.
- Demonstrate an ability to apply legal research results to solve legal problems.
- Developing a shared understanding of intellectual rigour

GA 12: Knowledge of a discipline

- Develop the capability of demonstrating comprehensive and considered knowledge of the discipline.
- Enables students to evaluate and utilize information and apply their disciplinary knowledge and their professional skills in the workplace.

3. Programme Specific Qualification Attributes

Mention the programme specific qualification attributes achieved through courses in the programme in terms of

- **Knowledge and understanding level (K1 and K2)**
  - Remember or recognize a term or a basic concept
  - Select an explanation for a statement related to the question topic
  - Understand the existing problems

- **Application-level (K3)**
  - Be able to solve the problems using computing techniques.

- **Analytical level (K4)**
  - Be able to separate information related to a procedure or technique into its constituent parts for better understanding and can distinguish between facts and inferences.

- **Evaluation capability level (K5)**
  - Be able to make judgments based on criteria and standards. Detects inconsistencies or fallacies within a process or product, determines whether a process or product has internal consistency and detects the effectiveness of a procedure as it is being implemented.

- **Scientific or synthesis level (K6)**
  - A scientific way to analyze and solve the problems.

4. Vision

Achieving excellence in Information Technology Enabled Services through Teaching, Research, Extension, and Consultancy.

5. Programme Objectives and Outcomes

Spelt the PEOs (Programme Educational Objectives), Programme Specific Objectives (PSOs) and Programme Outcomes (POs)

Programme Educational Outcomes (PEOs)

PEO1: Apply algorithmic, mathematical and scientific reasoning to a variety of computational problems

PEO2: Implement software systems that meet specified design and performance requirements.

PEO3: Work effectively in teams to design and implement solutions to computational problems

PEO4: Communicate effectively, both orally and in writing. Design, correctly implement and document solutions to significant computational problems

Programme Specific Outcomes (PSOs)

PSO1: An ability to apply profound knowledge to analyze and design software and systems containing hardware and software components of varying complexity.

PSO2: An ability to apply mathematical model, algorithmic principles, and computer science theory in the design of real-time applications

Programme Outcomes (POs)

PO1: Gain the skills and knowledge required to operate your computer and perform common tasks. This basic computer skills course will allow you to gain an understanding of the most popular, current technologies used at home and in the workplace.

PO2: The objective was set as training the students to use mathematical knowledge they have to designing and conducting 'Development Experiments' as well as analyzing and interpreting data using the Scientific Method.

PO3: Engage in effective software development practices over the entire system lifecycle including requirements, analysis, design, implementation, and testing.

PO4: Apply problem-solving skills and the knowledge of computer science to solve real problems.

PO5: Apply algorithmic reasoning to a variety of computational problems. Design, correctly implement and document solutions to significant computational problems

PO6: Learn new Programming concepts and coding skills are now essential for the high-tech industry. Everything is about new ways to build and use products and services that are more intuitive.

PO7: Theoretical knowledge and practical knowledge can often lead to a deeper understanding of a concept through seeing it in the context of a greater whole and understanding the why behind it.

PO8: Students will develop *critical* thinking skills. Students will develop an understanding of change *processes* and be able to think critically about obstacles to change.

PO9: Seeks and Progresses opportunities for change and growth. Is flexible and able to adapt to rapidly changing environments

PO10: It is a critical field that promotes systematic ways to design, evaluate, and manage computing solutions.

PO11: Intellectual rigour is encouraged for example during an assessment exercise where a debate or discussion occurs about a challenging topic.

PO12: Knowledge of a discipline is defined as "command of a discipline to enable a smooth transition and contribution to professional and community settings.

PEO-POMapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
PEO1												
PEO2												
PEO3												
PEO4												

PO-GAMAPPING:

	GA1	GA2	GA3	GA4	GA5	GA6	GA7	GA8	GA9	GA10	GA11	GA12
PO1												
PO2												
PO3												
PO4												
PO5												
PO6												
PO7												
PO8												
PO9												
PO10												
PO11												
PO12												

6. Candidate's eligibility for admission

A candidate who has passed Bachelors Degree in Electronics and Electrical Engineering/Electronics and communication Engineering/Computer Science and Engineering/Information Technology/Electronics and Instrumentation Engineering/Computer Science/Computer Technology/Software Engineering/BCA/ Mathematics/Statistics/Physics/Data Science/Data Analytics/ Electronics/Electronics and Communications under 10+2/3+2/3/4 system of this University or any of the degree of any other University accepted by the syndicate as equivalent thereto subject to such conditions as may be prescribed therefore shall be permitted to appear and qualify for the **M.Sc. Data Science** degree examination of this University after a course of study of two academic years.

7. Duration of the programme

The programme for the degree of **Master of Science in DATA SCIENCE** shall consist of **two Academic years** divided into four semesters. Each semester consists of 90 working days.

8. CBCS- Structure of the Programme

The programme structure comprises as follows.

Course Component	No. of Courses	No. of Courses * Hours of Learning/ Week	Marks	Credits
Part A (Credit Courses)				
Core Courses	10	40	1000	40
Elective Courses	03	12	500	12
Supportive Courses	1	3	100	3
Core - Practical	9	34	700	16
Project	1	-	200	20
Online Courses	3	3		
Total	27		2500	**91**

9. Curriculum structure for each semester as per your courses alignment

Course	L T P	No. of Hours / Week	Number of Credits
Semester-I			
CoreCourse- 20UPCSC4C01 Statistics and Probability	3+1+0	4	4
Core Course- 20UPCSC4C02 Mathematical methods for DataScience	3+1+0	4	4
Core Course- 20UPCSC4C03 Design of Algorithms	3+1+0	4	4
Core Course- 20UPCSC4C04 Bi Data Technologies	3+1+0	4	4
Core Course-20UPCSC4C05 Research Methodology	3+1+0	4	4
Core Course-20UPCSC4C06 R Programming – Lab	0+0+4	4	2
Core Course-20UPCSC4C07 SQL and Hadoop– Lab	0+0+4	4	2
Core Course- 20UPCSC4C08 Python Programming - Lab	0+0+2	2	1
SWAYAM / MOOC /SOFT SKILL-01 (Optional)	0+0+0		
Total		30	25

Semester-II			
Core Course- 20UPCSC4C09 Principles and Techniques of Data Science	3+1+0	4	4
Core Course- 20UPCSC4C10 Statistics Modeling	3+1+0	4	4
Core Course- 20UPCSC4C11 Machine Learning Techniques	3+1+0	4	4
Elective-I	3+1+0	4	4
Non-major Elective / Supportive Course	2+1+0	3	3
Core Course- 20UPCSC4C12 Machine Learning -Lab	0+0+4	4	2
Core Course- 20UPCSC4C13 Essential Technologies for DataScience(Weka& Intel AI software)	0+0+4	4	2
Core Course- 20UPCSC4C14 Mini Project/Summer Internship	0+0+2	2	1
SWAYAM / MOOC / SOFT SKILL-02 (Optional)	0+0+1	1	
Total		30	24

Semester-III			
Core Course- 20UPCSC4C15	3+1+0	4	4
Core Course- 20UPCSC4C16	3+1+0	4	4
Elective-II	3+1+0	4	4
Elective-III	3+1+0	4	4
Core Course- 20UPCSC4C17 Lab	0+0+4	4	2
Core Course- 20UPCSC4C18 Lab	0+0+4	4	2
Core Course- 20UPCSC4C19 Lab	0+0+4	4	2
SWAYAM / MOOC / SOFT SKILL-03 (Optional)	0+0+2	2	
Total		30	22

Semester-IV			
Option-I			
Core Course- 20UPCSC4C20 Internship based project. Dissertation and Viva Voice (Industry/Research)	20	-	20
Total			20
Total no of Credits **Core (10)**			40
Elective (3)			12
Practical (9)			16
Supportive (1)			3
Project (1)			20
Grand Total			91

Elective –I

20UPCSC4E01 Optimization Techniques

20UPCSC4E02 Advanced Statistical Techniques

20UPCSC4E03 Advanced Data Visualization

20UPCSC4E04 Discrete Mathematics

Elective –II

20UPCSC4E05

20UPCSC4E06

20UPCSC4E07

20UPCSC4E08

Elective – III

20UPCSC4E09

20UPCSC4E10

20UPCSC4E11

20UPCSC4E12

10. CBCS – Scheme of Examinations semester wise structure

Course	Number of Credits	Hours Per Week	Examination Duration (hrs)	Marks		
				I. A	ESE	Total
Semester-I						
CoreCourse- 20UPCSC4C01 Statistics and Probability	4	4	3	25	75	100
Core Course- 20UPCSC4C02 Mathematical methods for DataScience	4	4	3	25	75	100
Core Course- 20UPCSC4C03 Design of Algorithms	4	4	3	25	75	100
Core Course- 20UPCSC4C04 Bi Data Technologies	4	4	3	25	75	100
Core Course-20UPCSC4C05 Research Methodology	4	4	3	25	75	100
Core Course-20UPCSC4C06 R Programming – Lab	2	4	3	40	60	100
Core Course-20UPCSC4C07 SQL and Hadoop– Lab	2	4	3	40	60	100
Core Course- 20UPCSC4C08 Python Programming - Lab	1	2	3	40	60	100
SWAYAM / MOOC /SOFT SKILL-01 (Optional)						
Total	25					800
Semester-II						
Core Course- 20UPCSC4C09 Principles and Techniques of Data Science	4	4	3	25	75	100
Core Course- 20UPCSC4C10 Statistics Modeling	4	4	3	25	75	100
Core Course- 20UPCSC4C11 Machine Learning Techniques	4	4	3	25	75	100
Elective-I	4	4	3	25	75	100
Non-major Elective / Supportive Course	3	3	3	25	75	100
Core Course- 20UPCSC4C12 Machine Learning -Lab	2	4	3	25	75	100
Core Course- 20UPCSC4C13	2	4	3	40	60	100

Essential Technologies for DataScience						
Core Course- 20UPCSC4C14 Mini Project/Summer Internship	1	2	3	40	60	100
SWAYAM/MOOC/SOFT SKILL-02 (Optional)						
Total	24					800
Semester-III						
Core Course- 20UPCSC4C15	4	4	3	25	75	100
Core Course- 20UPCSC4C16	4	4	3	25	75	100
Elective-II	4	4	3	25	75	100
Elective-III	4	4	3	25	75	100
Core Course- 20UPCSC4C17 Lab	2	4	3	25	75	100
Core Course- 20UPCSC4C18 Lab	2	4	3	40	60	100
Core Course- 20UPCSC4C19 Lab	2	4	3	40	60	100
SWAYAM / MOOC / SOFT SKILL-03 (Optional)						
Total	22					700

Semester-IV						
Core Course-20UPCSC4C20 Dissertation and Viva Voice (Industry/Research)	20	--	3	50	150	200
Total	20					**200**
Total no of Credits: **Core** / **Elective** / **Practical** / **Supportive** / **Project**	40 12 16 3 20					
Grand Total	91					2500

EXAMINATIONS – THEORY

EVALUATION OF **INTERNAL ASSESSMENT**

Test	:	10 (5+5 Marks, 5 marks from best one of Test 1 and Test 2, 5 marks from test 3 –mode examinations)
Seminar	:	05Marks
Assignment	:	05Marks
Attendance	:	05Marks
Total	:	25Marks

EVALUATION OF **EXTERNAL EXAMINATIONS**

QUESTION PAPER PATTERN

Time duration: 3Hours Max. Marks:75

PART- A: 20× 1 = 20

Answer all the questions

(Objective type four questions from each unit)

PART- B: 3 × 5 = 15

Answer any three questions out of five questions

(Questions must be of type analytical)

PART- C: 5× 8 = 40

Answer all the questions

(Either or type for each unit)

The Passing minimum shall be 50% out of 75 marks (38 marks)

Part	**K Level**	**CO Coverage**
A: Objective type	K1, K2	CO1-20%, CO2-20%, CO3-20%, CO4-20% and CO5-20%
B: Analytical type	K3, K4, K5, K6	CO1-20%, CO2-20%, CO3-20%, CO4-20% and CO5-20%
C: Essay type	K1,K2, K3, K4, K5, K6	CO1-20%, CO2-20%, CO3-20%, CO4-20% and CO5-20%

PRACTICAL / SOFTWAREDEVELOPMENT

EVALUATION OF **INTERNALASSESSMENT**

Test 1	:	15Marks
Test 2	:	15Marks
Record	:	10 Marks
Total	:	40 Marks

EVALUATION OF **EXTERNAL EXAMINATIONS**

Time duration: 3Hours Max. Marks:60

QUESTION PAPER PATTERN

1. One compulsory question from the given listofobjectives : 30 Marks
2. One Either/OR type question from the given list of objectives : 30Marks

Distribution of Marks

ProblemUnderstanding	: 05Marks
Programwriting	: 10Marks
Debugging	: 10Marks
For CorrectResults	: 05 Marks

Mini-ProjectViva-Voce(joint): 60 Marks

DISSERTATION

Evaluation(External)	**: 50Marks**
Viva-voce(joint)	**: 100Marks**

11. REGULATIONS OF PROJECTWORK

a. Students should do their five months [Dec to Apr] Project work in Company /Institutions.
b. The Candidate should submit the filled-in format as given in Annexure-I to the department for approval during the Ist Week of January in their Projectsemester.
c. Each internal guide shall have a maximum of eightStudents.
d. Periodically the project should be reviewed a minimum three times by the advisorycommittee.
e. Students should prepare three copies of the dissertation and submit the same to the college on 30th April for the evaluation by examiners. After evaluation one copy is to be retained in the College Library and one copy is to be submitted to the University (Registrar) and the student can hold onecopy.
f. A Sample format of the dissertation is enclosed in Annexure-II.
g. The format of the Title page and certificate are enclosed in Annexure III.
h. The Students should use OHP / PowerPoint presentations during their Project Viva-voceExaminations.

12. Examinations

Examinations are conducted in semester pattern. The examination for the Semester I & III will be held in November/December and that for the Semester II and IV will be in the month of April/May. Candidates failing in any subject (theory, practical and skill) will be permitted to appear for such failed subjects in the same syllabus structure at subsequent examinations within next 3 years after normal completion of the programme. Failing which, the candidate has to complete the course in the present existing syllabus structure.

13. Scheme for Evaluation and Attainment Rubrics

Evaluation will be done on a continuous basis and will be evaluated four times during the course work. The first evaluation will be in the 7^{th} week, the second in the 11^{th} week, third in the 16^{th} week and the end – semester examination in the 19^{th} week. Evaluation may be by objective type questions, short answers, essays or a combination of these, but the end semester examination is a Universitytheory examination with prescribed question paper pattern.

ANNEXURE - I
PERIYAR UNIVERSITY

CollegeName :

Course :

StudentName :

RegisterNumber :

Title ofthe Project :

Address of Organization/Institution :

Name of the External Guide :

Designation :

Place :

Date: Signature of External Guide

(with seal)

Name of theInternal Guide :

Qualification :

TeachingExperience :

Place :

Date: Signature of Internal Guide

Principal [Approved or not Approved]
[University Use]

ANNEXURE II

COLLEGE BONAFIDE CERTIFICATE

COMPANY ATTENDANCE CERTIFICATE

Acknowledgment

CONTENTS

ANNEXURE III

Format of the titlepage

TITLE OF THE DISSERTATION

A Dissertation submitted in partial fulfillment of the requirements for the degree of

Master of Science in Data Science

to the

Periyar University, Salem - 11

By

STUDENT NAME

REG. NO.

College emblem

COLLEGE NAME (AFFILIATED TO PERIYAR UNIVERSITY)

PLACE with Pin Code

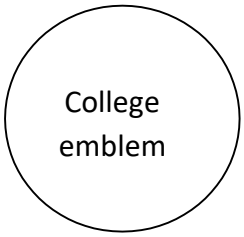
MONTH – YEAR

Format of theCertificate

COLLEGE NAME

(AFFILIATED TO PERIYAR UNIVERSITY)

PLACE with PIN CODE



MONTH – YEAR

PROJECT WORK

TITLE OF THE DISSERTATION

Bonafide Work Done

by STUDENT NAME

REG.NO

A Dissertation submitted in partial fulfillment of the requirements for the degree of **Master of Science in Data Science** to the **Periyar University, Salem -11.**

INTERNAL GUIDE HEAD OF THEDEPARTMENT

Submitted for the Viva-Voce Examination heldon _____

InternalExaminer External Examiner

M. Sc-DATASCIENCE - SYLLABUS

SEMESTER-I

COURSE CODE-20UPCSC4C01 **Credits: 4**

STATISTICS AND PROBABILITY

Course Objective:

- The main objective of this course is to provide students with the foundations of probabilistic and statistical analysis mostly used in varied applications in Business, Science etc.
- To use appropriate statistical methods to collect, organize, display, and analyze relevant data.
- To apply concepts of various probability distributions to find probabilities.

Unit 1: Descriptive Statistics- Different types of data – Measures of location, Measures of dispersion, Measures of shape – Empirical distribution

Unit 2: Statistical Charts and Diagrams: Histograms – Bar Charts – Frequency Curves – Density Plots – Box Plots – Bee Swarm Plots and Violin Plots – Drawing inferences from them

Unit 3: Bivariate Data – Correlations of various types – Partial and Multiple Correlations

Unit 4: Combinatorial probability – Counting Principle, Addition theorem – Conditional Probability – Multiplication Theorem – Bayes theorem

Unit 5: Random variables – Mean – Variance – Moments – Basic Discrete and Continuous distributions – Simulation from discrete and continuous distributions

Text Books:

1. David R. Anderson, Dennis J. Sweeney and Thomas A. Williams, Statistics for Business and Economics, 10th Edition, South-Western publication, 2008.

Book for Reference:

1. Ehsanes,A.K.Md. andRohatgi V.K.. An Introduction to Probability and Statistics, 2ndEdition, Wiley.
2. Gupta,S.C. and Kapoor, V.K., Fundamentals of Mathematical Statistics, Sultan Chand, 2017.

Course Outcomes

On the successful completion of the course, students will be able to

S.NO	Course Outcomes
CO1	Understanding Data, data attributes and describing it.
CO2	Understand the graphical methods of displaying data.
CO3	Observe and analyze the behavior of various discrete and continuous probability Distributions.
CO4	Understand the terminologies of basic probability, two types of random variables and their probability functions.
CO5	How to derive the probability density function of transformations of random variables and use these techniques to generate data from various distributions.

The mapping of course outcomes with programme outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	M	S	M	S	S	L	S	S	M	S	M	M
CO2	S	M	S	S	S	M	S	M	M	S	M	S
CO3	S	S	S	M	S	M	M	S	S	S	L	S
CO4	M	S	M	S	S	M	M	M	S	L	M	S
CO5	S	M	M	S	S	S	M	M	L	S	S	M

S- Strong; M-Medium; L-Low

COURSE CODE-20UPCSC4C02 **Credits: 4**

Mathematical methods for Data Science

Course Objective

- To provide suitable and effective methods called Numerical Methods, for obtaining approximate representative numerical results of the problems.
- To provide students with the foundations of essential linear algebra methods and Numerical methods to mostly use in varied applications in Business, Science etc.

Unit 1: Vectors & functions Systems of linear equations, Row reduction and echelon forms, Matrix operations

Unit 2: Linear dependence and independence, Subspaces and bases and dimensions, Orthogonal bases and orthogonal projections, Linear models and least-squares problems

Unit 3: Determinants and their properties, Inverse of a matrix, Cramer's Rule

Unit 4: Eigenvalues and eigenvectors, Positive definite matrices, Linear transformations, Singular Value Decomposition, Hessian matrix

Unit 5: Numerical Analysis: Iterative method, Taylor Series, Cauchy method, Newton Raphs Method.

Textbook:

1. Strang, Gilbert, et al. *Introduction to linear algebra*. Vol. 3. Wellesley, MA: Wellesley-Cambridge Press, Fifth Edition, 2016.
2. John B. Fraleigh, Raymond A. Beauregard, Victor J. Katz, *Linear algebra,* 3rd Edition,1995.

Reference Books:

1. Hoffmankunze, Introduction linear algebra, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 2nd endition,2009
2. David C. Lay, Steven R. Lay, Judi J. McDonald, Linear Algebra and Its Applications, Pearson publication,5th edition, 2018.

Course Outcomes

On the successful completion of the course, students will be able to

S.NO	Course Outcomes
CO1	Demonstrate understanding of basic mathematical concepts in data science, relating to linear algebra, probability, and calculus.
CO2	Employ methods related to these concepts in a variety of data science applications.
CO3	Analyze and evaluate the accuracy of common numerical methods.
CO4	Apply numerical methods to obtain approximate solutions to mathematical problems.
CO5	Apply Numerical analysis which has enormous application in the field of data Science

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	S	S	L	S	S	S	S	S	L	S	S	S
CO2	M	S	M	M	M	M	M	S	M	M	S	M
CO3	M	S	M	S	M	S	M	S	S	S	M	S
CO4	S	S	M	S	S	M	S	M	M	M	S	M
CO5	S	M	S	S	S	M	S	S	L	S	S	S

S- Strong; M-Medium; L-Low

COURSE CODE-20UPCSC4C03 **Credits: 4**

Design of Algorithms

Course Objective:

- To prepare a wiser consumer of data structures, algorithms, and heuristics.
- To explain different computational models (e.g., divideand-conquer), order notation and various complexity measures (e.g., running time, disk space) to analyse the complexity/performance of different algorithms.
- The analysis of the algorithm will be used to show the efficiency of the algorithm over the naive techniques.

Unit 1: Introduction: Algorithm, Psuedo code for expressing algorithms, Performance Analysis-Space complexity, Time complexity, Asymptotic Notation-Big oh notation, Omega notation, Theta notation and Little oh notation Elementary Data Structures: Stacks and Queues – Trees – Dictionaries –Priority Queues – Sets and Disjoint Set Union – Graphs

Unit 2: General Sorting method Bubble, Selection, Insertion, Divide and conquer- Merge & Quick sort, applications-Binary search, Quick sort, Merge sort, Linear Search

Unit 3: The Greedy Method: General Method - Container Loading - Knapsack Problem – Tree Vertex Splitting – Job Sequencing with Deadlines - Minimum Cost Spanning Trees – Optimal Storage On Tapes – Optimal Merge Patterns - Single Source Shortest Paths

Unit 4: Dynamic Programming: The General Method – Multistage Graphs – All-Pairs Shortest Paths – Single-Source Shortest Paths - Optimal Binary Search Trees - String Editing - 0/1 Knapsack - Reliability Design - The Traveling Salesperson Problem - Flow Shop Scheduling. Basic Traversal and Search Techniques: Techniques for Binary Trees – Techniques for Graphs – Connected Components and Spanning Trees – Biconnected Components and DFS

Unit 5: Backtracking: The General Method – The 8-Queens Problem – Sum of Subsets – Graph Coloring – Hamiltonian Cycles – Knapsack Problem Branch and Bound: Least Cost searched

Text Books:

1. Ellis Horowitz, SatrajSahni and SanguthevarRajasekaran, *Fundamentals of Computer Algorithms,* Universities Press, Second Edition, 2009

Reference Books:

1. T. H. Cormen, C. E. Leiserson, R. L. Rivest. *Introduction to Algorithms,* The MIT Press, Cambridge, Massachusetts, 3rd edition,2009.
2. Skiena, Steven S. *The algorithm design manual*. Springer International Publishing, 2020.

Course Outcomes

On the successful completion of the course, students will be able to

S.NO	Course Outcomes
CO1	Ability to analyze the performance of algorithms.
CO2	Demonstrate a familiarity with major algorithms and data structures.
CO3	Describe the Greedy Methods and explain when an algorithmic design situation calls for it. Synthesize greedy algorithms, and analyze them.
CO4	Understand the Dynamic Programming problems and their applications.
CO5	Synthesize efficient algorithms in common engineering design situations.

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	S	M	S	M	M	M	M	M	L	M	S	M
CO2	S	M	L	S	S	M	M	S	S	M	S	M
CO3	M	S	M	M	S	S	S	S	M	S	S	S
CO4	M	M	S	S	S	S	M	S	S	S	M	S
CO5	L	M	S	S	S	S	S	S	M	L	S	S

S- Strong ; M-Medium ; L-Low

COURSE CODE-20UPCSC4C04 **Credits: 4**

Big Data Technologies

Course Objective:

- To study the basic technologies that forms the foundations of Database.
- The course provides grounding in basic and advanced methods to big data technology and tools, including MapReduce and Hadoop and its ecosystem.
- To optimize business decisions and create competitive advantage with Big Data analytics

Unit 1: Introduction to Databases and Transactions What is database system, purpose of database system, view of data, relational databases, database architecture, transaction management, Database Design, ER-Diagram overview, ER-Model, Constraints, ER-Diagrams, ERD Issues, weak entity sets, Codd's rules, Relational Schemas, Introduction to UML Relational

Unit 2: Database model: Logical view of data, keys, integrity rules. Relational Database design: features of good relational database design, atomic domain and Normalization (1NF, 2NF, 3NF, BCNF).

SQL: data definition, aggregate function, Null Values, nested sub queries, Joined relations. Triggers, Transaction management: ACID properties, serializability and concurrency

Unit 3: Introduction to Big Data, World of Big Data: Why and Where, Characteristics of Big Data and Dimensions of Scalability, Foundations for Big Data Systems and Programming,

Unit 4: The Hadoop Ecosystem: Map Reduce, HDFS, Pig, Hive, Hbase, Scoop, yarn

Unit 5: NoSQL: Overview of NoSQL Databases Definition of the Four Types of NoSQL Database, The Value of Relational Databases, Getting at Persistent Data, Concurrency, Integration, Impedance Mismatch, Application and Integration Databases, Attack of the Clusters, The Emergence of NoSQL, Key Points, Comparison of relational databases to new NoSQL stores, MongoDB, Cassandra, HBASE, Neo4j use and deployment, Application, RDBMS approach, Challenges NoSQL approach, Key-Value and Document Data Models, Column-Family Stores, Aggregate-Oriented Databases

Text Books:

1. Silberschatz, Abraham, Henry F. Korth, and Shashank Sudarshan. *Database system concepts*. Vol. 4. New York: Mcgraw-hill, 1997.
2. Sadalage, Pramod J., and Martin Fowler. *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*. Pearson Education, 2013.
3. White, Tom. *Hadoop: The definitive guide*. " O'Reilly Media, Inc.", 2012.

Reference Books:

1. Kenneth Cukier and Viktor Mayer-Schönberger, Big Data: A Revolution That Will Transform How We Live, Work, and Think, Hodder And Stoughton, 2013.
2. Rob, Coronel, "Database Systems", Course Technology Inc, 7th edition, 2006.
3. Eric Redmond & Jim R. Wilson , Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement, O'Reilly publisher, 2012.

Course Outcomes

On the successful completion of the course, students will be able to

S.NO	Course Outcomes
CO1	Illustrate the usage of data on different data ecosystems.
CO2	To design and build a simple database system and demonstrate competence with the fundamental tasks involved with modelling, designing, and implementing a DBMS.
CO3	Understand the fundamentals of various big data techniques.
CO4	Analyze the HADOOP and Map Reduce technologies associated with big data analytics.
CO5	Develop an application using different eco system tools by taking standard sample data set.

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	S	S	M	S	M	M	S	M	M	L	S	S
CO2	S	M	S	S	S	M	M	S	M	S	S	M
CO3	M	M	M	S	S	S	M	S	S	M	S	S
CO4	S	S	M	S	M	L	M	S	M	S	M	S
CO5	M	S	M	M	S	S	M	S	S	L	M	S

S- Strong ; M-Medium ; L-Low

COURSE CODE-20UPCSC4C05 **Credits: 4**

Research Methodology

Course Objective:

- The primary objective of this course is to develop a research orientation among the students and to acquaint them with fundamentals of research process.
- To develop an understanding of various research designs and techniques and to identify various sources of information for literature review and data collection.
- To learn the art of writing the reports and to be cautious of plagiarism.

Unit 1: Objectives and types of research Definition of Research – Importance, limitations – Motivation and objectives – Research methods vs. Methodology. Types of research – Descriptive. Analytical, Applied vs. Fundamental, Quantitative vs. Qualitative, Conceptual vs. Empirical, concept of applied and basic research process, criteria of good research.

Unit 2: Research Formulations and Design Defining and formulating the research problem, selecting the problem, necessity of defining the problem, importance of literature review in defining a problem, literature review-primary and secondary sources, reviews, monograph, patents, research databases, web as a source, searching the web, critical literature review, identifying gap areas from literature and research database, development of working hypothesis.

Unit 3: Data Collection and Analysis Computer and its role in research, Effective use of Internet, Execution of the research – Observation and Collection of data – Methods of data collection

Unit 4: Reporting and thesis writing Structure and components of scientific reports –Types of report – Technical reports and thesis – Significance – Different steps in the preparation – Layout, structure

Unit 5: Language of typical reports – Illustrations and tables- Bibliography, referencing and footnotes – Oral presentation – Planning – Preparation –Practice – Making presentation – Use of visual aids – Importance of effective communication.

Text Books

1. John W. Best, James V. Kahn (2005) Research in Education. Tenth Edition, Pearson Publisher.

2. Kothari, C.R. (2019) Research Methodology Methods and Techniques. Fourth Edition, New Age International Publishers, New Delhi.

Reference Books

1. Fred N. Kerlinger (2017) Foundations of Behavioural Research. Surjeet Publications.

2. Rand R.wilcox. (2010) Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy, Second edition, Springer.

Course Outcomes

On the successful completion of the course, students will be able to

S.NO	**Course Outcomes**
CO1	Students would have understood the concepts of research process in science.
CO2	To design the Research Proposal and decide the sample techniques and size
CO3	Students would have become familiar with the mode of collecting data and do the interpretation of the same.
CO4	Independently work in a research environment, consolidate the outcome of research and write technical papers.
CO5	Have basic knowledge on how to present research techniques.

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	M	S	L	S	S	S	M	M	S	S	S	M
CO2	S	S	M	M	S	M	S	M	M	L	M	S
CO3	M	M	M	S	S	S	M	S	M	M	M	S
CO4	S	M	S	S	M	M	S	S	S	L	S	M
CO5	S	L	M	S	S	M	S	M	S	S	S	S

S- Strong; M-Medium ; L-Low

COURSE CODE-20UPCSC4C06 **Credits: 2**

R Programming Lab

Implement the following in R studio.

1 Installing R on Windows, and R Studio

2 Data Types - R Objects and Attributes, Data Types - Vectors and Lists

3 Data Types - Data Frames, Matrices, Factors & Functions

4 Expression & Logical Statement in R

5 Subsetting of List, Matrices, &Dataframe

6 Dataframe functions on inbuilt Dataset

7 Dplyr Function on Dataset

8 Dplyr Function on Dataset

9 Basic Plotting with R

10 GGplots with R

11 Working with R Markdown & pushing Code to Git

12 Building R Shiny Dashboard App

COURSE CODE-20UPCSC4C07 **Credits: 2**

SQL and Hadoop Lab

Implement the following in Hadoop.

1	Viewing all databases, Creating a Database, Viewing all Tables in a Database, Creating Tables (With and Without Constraints), Inserting/Updating/Deleting Records in a Table, Saving (Commit) and Undoing (rollback)
2	Altering a Table, Dropping/Truncating/Renaming Tables, Backing up / Restoring a Database
3	Real time Retail dataset Database creation
4	Queries with Aggregate functions (group by and having clause)
5	Data Analysis with SQL on Big Mart Dataset
6	Join Queries- Inner Join, Left & Right Join, Full Join, Cross Join
7	Startup VM, Hadoop 1.0 & Linux Command Basic
8	Transfer of file to HDFS & Map Reduce Program-I (Counting Problem)
9	Map Reduce Program-II (Data Analysis on Sales dataset)
10	Basic Hive Commands
11	Advance Hive Commands
12	Real time Data analysis with scoop & Hive

COURSE CODE-20UPCSC4C08 **Credits: 1**

Python Programming Lab

Implement the following in Python.

1	Introduction to Python, How to Use Shell, SpyderandJupyter Notebook
2	Values, Data Types, Operators & Function Calls
3	Type Casting, Statements and Expressions
4	Input & Output Functions, Pythonic Functions, Exception Handling
5	Python OOPS Basic
6	Python OOPS Advance
7	Python GUI Development
8	Python Module: OS, Numpy
9	Python Module: Pandas
10	Python Module: Matplotlib and Seaborn
11	Python Module: Scipy
12	Python Module: Sklearn Pre-processing and Imputation

COURSE CODE-20UPCSC4C09 **Credits: 4**

Principles & Techniques of Data Science

Course Objective:

- To identify the scope and essentiality of Data warehousing and Data Mining.
- To develop research interest towards advances in data mining.
- To analyze the data, data science lifecycle, data collection and cleaning, exploratory data analysis and visualization, statistical inference and prediction, and decision-making algorithms for respective applications.

Unit 1: Why Data Mining? Moving toward the Information Age Data Mining as the Evolution of Information Technology, What Is Data Mining, What Kinds of Data Can Be Mined, Database Data, Data Warehouses, Transactional Data, Other Kinds of Data, OLTP & Online Analytical Processing(OLAP), Graphs Database

Unit 2: Getting to Know Your Data: Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization, Measuring Data Similarity and Dissimilarity: Euclidean, Jaccard's Index & Cosine Similarity

Unit 3: Data Pre-processing on Big Data: Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization (ETL Operations)

Unit 4: Data Science Tools: One API AI toolkit

Unit 5: Data Science Applications in Uses Cases BSFI, Retail, Telecom & Healthcare

Text Books:

1. Daimi, Kevin, et al. *Principles of Data Science*. Ed. Hamid R. Arabnia. Springer, 2020.
2. Han, Jiawei, Micheline Kamber, and Jian Pei. "Data mining concepts and techniques third edition." *The Morgan Kaufmann Series in Data Management Systems* 5.4 (2011).
3. Boston Provided Material for (Unit 4 & 5)

Reference Books

1. SinanOzdemir Principles of Data Science: Mathematical techniques and theory to succeed in data-driven industries, Packt Publishing Limited (13 December 2016)

2. Cielen, Davy, Arno DB Meysman, and Mohamed Ali. *Introducing data science: big data, machine learning, and more, using Python tools*. Manning Publications Co., 2016.

Course Outcomes

On the successful completion of the course, students will be able to

S.NO	Course Outcomes
CO1	Understand fundamentals of data and Data Mining Principles.
CO2	To Understand importance of qualitative data, terminologies related to Data Science.
CO3	Understand and Extract knowledge using data processing concepts in data science.
CO4	Evaluate the databases, file formats and its applications using API Toolkit.
CO5	Analyze and design data mining applications.

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	M	S	L	S	M	M	L	S	L	S	M	M
CO2	M	S	M	M	S	S	S	M	M	M	M	S
CO3	M	M	S	S	M	S	M	S	M	L	M	S
CO4	S	S	M	S	S	S	M	S	MZ	S	S	S
CO5	S	M	S	M	M	S	M	M	S	M	M	S

S- Strong; M-Medium ; L-Low

COURSE CODE-20UPCSC4C10 **Credits: 4**

Statistical Modeling

Course Objective:

- To understand the hypothesis testing and able to calculate confidence interval.
- To study the introduction to the principles of statistical inference and linear statistical models.
- To learn about the central tendency, correlation and correlation coefficient and also regression and interpret the results.
- To understand the probability density function of transformations of random variables and use these techniques to generate data from various distributions.
- To know the non-parametric methods such as Kolmogrovsmirnov tests – Kruskal Wallis Test.

Unit 1 : Introduction to Statistical Inference : Estimation – Testing of Hypothesis – Confidence Intervals – Sampling Distributions

Unit 2 : Tests of significance related to hypothesis about mean, variance – testing of equality of means- confidence intervals based on statistical testing

Unit 3 : Simple and multiple linear regression models – estimation, testing and confidence intervals of parameters in regression models – Prediction – R-Square and adjusted R-square

Unit 4 : Analysis of Variance – One way and Two-Way analysis – Multiple comparison tests

Unit 5 : Non-parametric methods: Sign and Median tests – Kolmogrovsmirnov tests – Kruskal Wallis Test

Text Books:

1.Anderson, D.R., Sweeney, D.J. and Williams, T.A., "Statistics for Business and Economics", 10th Edition, Thompon South-Western, 2008

Book for Reference:

1.Ehsanes, A.K.Md. and Rohatgi V.K., "An Introduction to Probability and Statistics", 2nd Edition, Wiley,2005

2.Gupta, S.C and Kapoor, V.K., "Fundamentals of Applied Statistics", Sultan Chand, 2018.

Course Outcomes

On the successful completion of the course, students will be able to

S. No	**Couse Outcome**
CO1	Analysis of the complex problems arising in real life problems.
CO2	To Observe and analyze the behavior of various discrete and continuous probability distributions.
CO3	To interpret all of the results of Regression models and the Correlation Analysis.
CO4	To apply discrete and continuous probability distributions to various problems in daily life.
CO5	To derive the probability density function of transformations of random variables the techniques to generate data from various distributions.

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	M	L	M	S	M	S	M	S	M	L	S	M
CO2	M	M	S	S	S	M	M	L	M	M	S	S
CO3	S	M	M	M	S	M	S	S	M	S	S	S
CO4	M	S	M	M	M	M	M	S	L	M	M	S
CO5	M	S	S	S	L	S	S	M	S	S	M	S

S- Strong ; M-Medium ; L-Low

COURSE CODE-20UPCSC4C11 **Credits: 4**

Machine Learning Techniques

Course Objective:

- To introduce students to the basic concepts and techniques of Machine Learning.
- To understand the regression methods, regularization methods, clustering methods and metrics.
- To learn Decision trees, CART and Ensemble Techniques.
- To understand the problems using various machine learning techniques
- To study the recent machine learning software for solving practical problems.

Unit 1: Introduction of Machine Learning, Learning problem & System, Designing a learning system, concept of learning task, Hypothesis Space & Version Space

Unit 2: Introduction to Machine Linear Regression, Machine Logistics Regression, Regularization, Cross Validation & Metrics

Unit 3: Introduction to Decision Tree, CART, & Ensemble Techniques

Unit 4: Introduction to SVM, Kernels, Neural Networks (MLP), Collaborative filtering

Unit 5: Clustering Algorithms: K Means, K Mode, K Median,Hierarchical clustering, DBSCAN, PCA, TSNE

Text Books:

1. Tom M. Mitchell, *"Machine Learning*", McGraw Hill,U.K, 2017.

2. Shalev-Shwartz, Shai, and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

Reference Books

1. Saleh, Hyatt. *Machine Learning Fundamentals: Use Python and scikit-learn to get up and running with the hottest developments in machine learning*. Packt Publishing Ltd, 2018.

Course Outcomes

On the successful completion of the course, students will be able to

S. No	Couse Outcome
CO1	Apply the machine learning concepts in real life problems.
CO2	To Implement and analyze existing learning algorithms, including well-studied methods for classification, regression, clustering.
CO3	To Identify machine learning techniques suitable for a given problem.
CO4	To Design application using machine learning techniques.
CO5	To Solve the problems using various machine learning techniques

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	M	S	S	S	S	M	S	S	M	M	S	S
CO2	M	M	S	M	S	M	M	S	M	M	S	S
CO3	M	S	L	S	S	S	S	M	M	M	L	M
CO4	S	S	M	S	S	M	S	M	S	S	M	M
CO5	S	M	M	M	M	L	M	S	M	M	S	S

S- Strong; M-Medium; L-Low

COURSE CODE-20UPCSC4C12 **Credits: 2**

Machine Learning Lab

Implement the following.

1	Introduction to Python Framework on ML & significations
2	Linear Regression on Real Time Dataset
3	Logistics Regression on Real Time Dataset
4	Lasso & Ridge Regression on Real Time Dataset
5	KNN on Real Time Dataset
6	Decision Tree on Real Time Dataset
7	SVM on Real Time Dataset
8	MLP on Real Time Dataset
9	Random Forest on Real Time Dataset
10	K-Means on Real Time Dataset
11	DBSCAN on Real Time Dataset
12	TSNE & PCA on Real Time Dataset

COURSE CODE-20UPCSC4C13 **Credits: 2**

Essential Technologies for Data Science Lab

Implement the following.

1 Introduction to Intel NumPy, Pandas & DAAL

2 Introduction to MKL

3 Data Engineering on Banking Dataset

4 Data Analysis on Banking Dataset

5 Building Dashboard on Sales dataset

6 Data Engineering on Banking Dataset

7 Data Analysis on Banking Dataset

8 Clustering on Sales dataset

9 Time series Data Understanding & Handling

10 Trend Analysis

11 Smoothing, Lag & Pre-processing of Time series Data

12 Forecasting Models with Time series Models, AR, MA, ARIMA

COURSE CODE-20UPCSC4C14 **Credits: 1**

Mini Project and Summer Internship

Guidelines

Problem Statement on any topic from semester 2 subjects

Survey and study of published literature on the assigned topic; Working out a preliminary Approach to the Problem relating to the assigned topic; Conducting preliminary Analysis/Modelling/Simulation/Experiment/Design/Feasibility; Preparing a Written Report on the Study conducted for presentation to the Department; Final Seminar, as oral Presentation before a Departmental Committee.

COURSE CODE-20UPCSC4E01 **Credits: 4**

Optimization Techniques

Course Objective:

- To understand the fundamental knowledge of Linear Programming and Dynamic Programming problems.
- To Learn the classical optimization techniques and numerical methods of optimization.
- To study the basics of different evolutionary algorithms.
- To understand the Integer programming techniques and apply different optimization techniques
- To solve various models arising from engineering areas.

Unit 1: Linear Programming Problem Linear Programming Problem (LPP): Mathematical Formulation of Linear Programming Problem - Graphical Solution of LPP - canonical and standard forms of linear programming problem- Simplex method for solving LPP

Unit 2: Transportation Model: North West corner Method, Least cost method, and Vogel's Approximation Method. Assignment Model: Hungarian assignment model – Travelling Sales Man Problem

Unit 3: Project Scheduling PERT/CPM Networks – Fulkerson's Rule – Measure of Activity – PERT Computation – CPM Computation – Resource Scheduling.

Unit 4: Simplex Method – Gradient of function – Steepest Descent method – Conjugate Gradient method

Unit 5: Particle Swarm Optimization method – Ant Colony optimization algorithm – Fruit Fly method – Fire Fly method

Text Books:

1. Christos H. Papadimitriou and Kenneth Steiglitz, "Combinatorial Optimization -- Algorithms and Complexity", Dover Publications, 2013

2. David G. Luenberger and Yinyu Ye, "Linear and Non-linear Programming", Stanford University, Springer,4TH Edition, 2016.

3. Stephen Boyd and Lieven Vandenberghe, "Convex Optimization", Cambridge University Press, 2004.

Reference Books

1. Hamdy A. Taha, "Operations Research: An Introduction", Pearson, 2010

Course Outcomes

On the successful completion of the course, students will be able to

S. No	Course Outcomes
CO1	To develop the fundamental Linear Programming and Dynamic Programming problems of industrial process.
CO2	Learn efficient computational procedures to solve optimization problems.
CO3	To apply the basics of different evolutionary algorithms.
CO4	To analyze the fundamentals of Integer programming technique.
CO5	To apply different techniques to solve various optimization problems arising from business areas.

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	S	L	S	S	S	S	L	S	M	S	L	M
CO2	S	M	S	L	S	M	M	M	S	S	M	S
CO3	M	S	S	S	S	M	S	S	M	M	M	S
CO4	S	M	M	M	M	S	S	M	S	M	S	S
CO5	M	M	M	S	S	M	M	M	M	M	M	S

S- Strong; M-Medium; L-Low

COURSE CODE-20UPCSC4E02 **Credits: 4**

Advanced Statistical Techniques

Course Objective:

- To understand the fundamental knowledge of Multivariate distributions, variance, covariance matrix.
- To Learn the Linear Regression Analysis and Ridge Regression.
- To study the basics of PCA and SVD algorithms.
- To understand the ECM, Linear Discriminant analysis and Logistic regression
- To solve ARIMA models for time series forecasting models.

Unit 1 (Multivariate Analysis): Notion of random vector, Mean vector , Variance-Covariance matrix and its properties – Multivariate normal distribution – sample generation from normal distribution – Computing partial and multiple correlations from variance- covariance matrix.

Unit 2 Multiple Linear Regression Analysis – Estimation and Testing – Multicollinearity – Ridge Regression – Variable selection

Unit 3 Principal Component Analysis – Interpreting PCAs – its usage- Singular Value Decomposition – Applications in Data Science

Unit 4 Classification: Minimum ECM rule – Mahalanobis D-square – Linear Discriminant analysis for two classes and more than classes problems – performance measures – Logistic regression

Unit 5 Time Series Models: Auto-regressive models – Moving average models – ARIMA models – Forecasting using time series models.

Text Books and Reference:

1. Makridakis, S, Wheelwright, S, C. and Hyndman, R.J., "Forecasting: Methods and Applications", 3rd Edition, Wiley Publications(Unit 1,3,4).

2. Montgomery, D.C., Peck, E.A. and Vining, G.G., "Introduction to Linear Regression Analysis", 5th Edition, Wiley Publications. (Unit 2)

3. Johnson, R.A and Wichern, D.W, "Applied Multivariate Statistical Analysis", 6th Edition, Pearson Publications. (Unit 5)

Course Outcomes

On the successful completion of the course, students will be able to

S. No	Course Outcome
CO1	To develop multivariate distribution and their applications in data science.
CO2	To use regression techniques and variable selection methods.
CO3	Implement dimension reduction techniques using software on real life problems.
CO4	To Analyze performance measure of LDA and regression technique
CO5	To apply different techniques to solve various time series forecasting problems arising from engineering areas.

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	S	S	M	S	S	L	S	S	S	M	S	S
CO2	S	S	S	M	M	M	S	S	M	S	M	M
CO3	S	M	S	S	S	S	M	S	M	S	M	M
CO4	M	M	M	S	S	S	M	S	M	M	S	M
CO5	M	M	M	S	S	M	S	S	M	M	M	S

S- Strong ; M-Medium ; L-Low

COURSE CODE-20UPCSC4E03 **Credits: 4**

Advanced Data Visualization

Course Objective:

- To introduce students to the basic concepts and techniques of data visualization.
- To design and visualize various data including multivariate data, networks, text and cartography.
- An understanding of the key techniques and theory used in visualization, including data models, graphical perception and techniques for visual encoding and interaction.

Unit 1 Introduction : The case for data visualization and storytelling, Data visualization and storytelling details and best practices, Introduction to Tableau, PowerBI Importing Data / Connecting to External Sources Interface Overview Creating Sheets and Dashboard (software Demo)

Unit 2 Elements of Data Visualization: Measures & Dimension concepts, Design Fundamentals Design Principles, Colors, and "Chart Junk" Design perspectives from the experts The Shaffer 4 C's of Data Visualization Not-so-best practices (examples) Critique and redesign, Creating a good data set for analysis Data modeling fundamentals for analytics Selecting data for your KPIs,

Unit 3 Story telling with Data: Depth Design Fundamentals of Data Visulization, What are the main approaches to storytelling with data? - Dashboards vs. Storyboards vs. Info graphics - Designing with the user in mind The Duell Rules for Actionable Visualizations, Basic Charting Techniques (Hands-on Session)

Unit 4 Interactive Elements & Advance Graphics : Interactive Visualization Features – build interactive visualization Actions and filters Calculated measures Data blending, joins, and custom queries Custom Shape Files, Multiple choice and short answer on Blackboard Create graphs appropriate for data Chart critique Geocoding and Mapping, Advanced Chart types Custom Color Palettes

Unit 5 Case Studies: Compare and Contrast real-world examples Flowing Data - Nathan Yau Information is Beautiful Tableau Vizzes in the wild

Textbook

1. Yau, Nathan. *Visualize this: the FlowingData guide to design, visualization, and statistics*. John Wiley & Sons, First Edition, 2011.

2. Few, Stephen. *Information Dashboard Design: Displaying data for at-a-glance monitoring*. Vol. 5. Burlingame, CA: Analytics Press, Second Edition, 2013.

Reference Books

1. Hamilton, Max. "The visual display of quantitative information. Edward R. Tufte. Graphics Press, Cheshire, Connecticut, 5th printing (2001).
2. Knaflic, Cole Nussbaumer. *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, First Edition, 2015.
3. Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. *The big book of dashboards: visualizing your data using real-world business scenarios*. John Wiley & Sons, First Edition, 2017.

Course Outcomes

On the successful completion of the course, students will be able to

S. No	Course Outcome
CO1	To demonstrate proficiency with statistical analysis of data.
CO2	To introduce the tools required to manage and analyze the elements of data visualization.
CO3	Apply visual design principles to simple and complex models that tell the stories found in data.
CO4	To understand how to work with data for real life.
CO5	To apply mathematical and statistical models and concepts to detect patterns in data, as well as draw inferences and conclusions supported by the data.

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	S	S	S	S	M	S	S	S	S	M	S	M
CO2	M	S	S	S	M	M	S	M	S	S	M	S
CO3	M	M	M	M	S	M	S	L	M	S	M	S
CO4	M	S	M	S	S	M	M	M	M	S	M	M
CO5	S	S	L	S	S	M	M	S	M	S	S	S

S- Strong ; M-Medium ; L-Low

COURSE CODE-20UPCSC4E04 **Credits: 4**

Discrete Mathematics

(Theorems and Proofs are not expected)

Objectives:

- To understand the applications of functions and relations
- To understand the basic concepts of mathematical logic and predicate calculus
- To understand the concept of the method of induction
- To develop the skills in solving recurrence relations

Unit-I

Well-formed formulas – truth table of well-formed formula – a tautology, contradiction, and contingency –the equivalence of formulas. Algebra of propositions – Functionality complete sets – Normal forms of well-formed formulas- Rules of Inference for propositional calculus – well-formed formulas of predicate calculus – Rules of Inference for predicate calculus – predicate formulas involving two or more quantifiers. (2.1-2.11)

Unit-II

Set theory – relations- functions – set identities – Binary relations – properties of binary relations in a set – Equivalence relations and partial orderings – Representation of a relation by a matrixrepresentation of a relation by a digraph - Basics of Counting – Integers and Induction. (1.3-1.7, 4.1-4.2, 5.1-5.5)

Unit-III

Formulation as Recurrence Relations-solving recurrence Relation by Iteration-solving Recurrence Relations- Solving Linear Homogeneous Recurrence Relations of Order Two-Solving Linear Nonhomogeneous Recurrence Relations. Permutations-Combinations-Permutations with repetitions-Combinations with repetition-permutations of sets with indistinguishable objects.(6.1-6.5, 3.1-3.6)

Unit-IV

Definition and examples-properties of lattices –lattices as algebraic systems-Sub lattices and lattice Isomorphism-special classes of lattice –distributive lattices and Boolean algebras.(8.1-8.6)

Unit-V

Connected Graphs-Euler Graphs-Hamiltonian circuits and paths – planar graphs – matrix representation of graphs. (10.1-10.5 and 10.8)

Text Book:

1. N.Chandrasekaran and M.Umaparvathi, Discrete mathematics, PHI Learning Private Limited, New Delhi, 2010

References:

1. J.P.Trembley and R.Manohar, *Discrete Mathematical Structures with Applications to Computer Science,* Tata McGraw Hill, New Delhi, 2001.
2. Sengadir, T. *Discrete mathematics and combinatorics*. Pearson Education India, 2009.
3. RakeshDube, AdeshPandey, Ritu Gupta, *Discrete Structures and Automata Theory*, Narosa Publishing House, New Delhi, First Edition, 2006.

Course Outcomes

On the successful completion of the course, students will be able to

S. No	Course Outcomes
CO1	Understand the use of well-formed formula, logical connectives, predicates of formula and so on.
CO2	Evaluate Relations, functions, recurrence relations and their types.
CO3	Apply lattices with use of Hassee diagram and study of its properties and various study of different graphs.
CO4	Demonstrate an understanding of relations and functions and be able to determine their properties.
CO5	To verify whether an algorithm works well and perform analysis in terms of memory and time.

The mapping of course outcomes with program me outcomes is tabulated as follows.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	S	S	M	S	M	L	M	S	M	M	M	S
CO2	S	S	M	S	M	M	L	S	L	S	S	M
CO3	S	M	S	S	M	L	M	M	S	S	M	S
CO4	M	M	M	M	S	M	S	M	M	S	M	M
CO5	S	S	M	M	S	M	M	M	S	S	M	S

S- Strong ; M-Medium ; L-Low