SIES College of Arts, Science & Commerce
RISE WITH EDUCATION
Sion (West), Mumbai – 400022.
(Autonomous)

Faculty: Science
Program: B.Sc.
Subject: Data Science
Academic year: 2023-2024

Syllabus for Semester- III and Semester – IV

Credit Based Semester and Grading Syllabi approved by Board of Studies in Data Science to be brought into effect from June 2023.

Preamble

Data has become the most important factor in this era of digital transformation. The technological advancements are seen in all walks of life and therefore we are flooded with enormous data. Every business relies on data to deliver better products as well as services. All data are stored in cloud, and so accessed and processed easily. Data analytics has helped in better decision making with sufficient data insights.

Predictive Analysis has played a crucial role in making businesses smarter with improvised strategies. Machine Learning and Artificial Intelligence are used together to optimize businessoperations and data management. Augmented analytics uses machine learning and natural language processing to automate the process of data analysis. Global data is predicted to grow due to data generated by the Internet of Things (IoT) and cloud computing advancements. Thesedevelopments have given rise to a new area of study, called Data Science.

Data Science as an area has evolved out of the applications of various tools and techniques inthe field of Computer Science, Mathematics and Statistics. There is an increasing demand tocapture, analyse the enormous data present in a number of application domains. The data in these applications then needs to be converted into actionable strategies for effective decisionmaking. So, the study of data science has become essential to cater to the growing need of data scientists and data analysts.

This course focuses on educating the students about the essentials of computer science,applied mathematics and applied statistics with respect to the data science applications.

PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES

NO.	Details
PO1.	Solving Complex Problems :- Apply the knowledge gained in breaking down complex problems into simple components; and to design processes required for problem solving.
PO2.	Critical Thinking: - Ability to apply the acquired knowledge to identify assumptions and evaluate their accuracy and validity.
PO3.	Reasoning ability and Rational thinking: - Ability to analyse, interpret data and draw logical conclusions; to evaluate ideas rationally.
PO4.	Research Aptitude: - Ability to ask relevant questions to identify and define the problem, applying research tools for analysis and interpretation of data. Understand and comply with research ethics.
PO5.	Information and Digital Literacy: - Equip to use appropriate tools and techniques inclusive of internet and electronic media for acquiring, assessing and analysing data from diverse resources.
PO6.	Social Interactive Skills and team work: - Exhibit networking and social interactiveskills; function effectively as an individual and as a member in diverse groups; demonstrate leadership quality useful for employability
PO7.	Self-directed and Lifelong Learning: Ability to explore and gain knowledge in independent and self-reliant ways. Demonstrate ability to adapt and upgrade with the global , social and technological changes.
PSO1.	Sound Knowledge: Demonstrate the knowledge of core data science concepts and apply them to develop a user- friendly, scalable and robust applications
PSO2.	Critical and Rational Thinking: Exhibit higher order skills to adapt to the everchanging technological environment
PSO3.	Logic Building and Programming Skills: The ability to apply logic to problem solving and acquire proficiency in various programming languages.
PSO4.	Data Analysis : Apply quantitative modeling and data analysis techniques to solve real world business problems, Learn tools and techniques for transformation of data and statistical data analysis
PSO5.	Pursue Higher Education: Make students competent to take up advanced degree courses like MSc(Data Science),MCA, MSc(CS), MSc(IT) and MBA etc.

SYBSc(DS) - Semester III

Course Code	Course Type	Course Title	Periods per week (1 Period = 50 min)	Credits
SIUSDS31	Core Subject	Web Technology	5	2
SIUSDS32	Core Subject	Data Structures and Program Design Using Python	5	2
SIUSDS33	Skill Enhancement Course	Probability and Distributions	5	2
SIUSDS34	Core Subject	Testing of Hypothesis	5	2
SIUSDS35	Core Subject	Introduction to Business Analytics	5	2
	Core Subject			
SIUSDSP31	Core Subject Practical	Web Technology Practical	3	2
SIUSDSP32	Core Subject Practical	Data Structures and Program Design Using Python Practical	3	2
SIUSDSP33	Skill Enhancement Course Practical	Probability and Distributions Practical	3	2
SIUSDSP34	Core Subject Practical	Testing of Hypothesis Practical	3	2
SIUSDSP35	Core Subject Practical	Introduction to Business Analytics Practical	3	2
		TOTAL CREDITS		20

SEMESTER – 3

SYBSc(DS) - Semester III

Web Technology (SIUSDS31)

Course Objective:

- Introducing the basic concepts of Internet and web design to learners
- Providing brief knowledge about HTML5 concepts
- Giving insight of the Page layout and navigation with HTML5
- Making students aware about use of Tables, Forms and Media with HTML5
- Teaching data validation using Java Script. Giving knowledge about transmission of data on web page using JSON object.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Understand the meaning of the basic terminologies of web technology and explore, use The HTML5 concepts, Understand the basic requirement of web design.
CO2: Understand and use the Page layout, Navigation, Tables, Forms and Media features of HTML5.
CO3: Understand and use Cascading Style sheet for beatifying the webpages.
CO4: Understand and use the Java Script for validation of user forms in web pages.
CO5: Understand and use the technique of transmitting data between a server and web application using JSON

Unit	Contents	No. of Lectures
I	**Internet and the World Wide Web:** What is Internet? Introduction to internet and its applications, E-mail, telnet, FTP, e-commerce, video conferencing, e-business. Internet service providers, domain name server, internet address **World Wide Web (WWW):** World Wide Web and its evolution, uniform resource locator (URL), browsers – internet explorer, Netscape navigator, opera, Firefox, chrome, Mozilla. search engine, web saver – apache, IIS, proxy server, HTTP protocol **What Is Web Design?:** Defining Web Design, Web Design Themes, Learning Web Design. **User-Centered Design:** Usability, Who Are Web Users? Common User Characteristics, Memory, Response and Reaction Times, Dealing with Stimulus, Movement Capabilities, The User's World, General Types of Users, Web Conventions, Accessibility, Building a Usable Site **HTML5:** Introduction, Why HTML5? Formatting text by using tags, using lists and backgrounds, Creating hyperlinks and anchors.	12

II	**HTML5 Page layout and navigation:** Creating navigational aids: planning site organization, creating text based navigation bar, creating graphics based navigation bar, creating graphical navigation bar, creating image map, redirecting to another URL, creating division based layouts: HTML5 semantic tags, creating divisions, creating HTML5 semantic layout, positioning and formatting divisions. **HTML5 Tables, Forms and Media:** Creating tables: creating simple table, specifying the size of the table, specifying the width of the column, merging table cells, using tables for page layout, formatting tables: applying table borders, applying background and foreground fills, changing cell padding, spacing and alignment **Creating user Forms:** creating basic form, using check boxes and option buttons, creating lists, additional input types in HTML5, Incorporating sound and video: audio and video in HTML5, HTML multimedia basics, embedding video clips, incorporating audio on web page.	12
III	**Introduction to Style Sheets :** Understanding Styles, Constructing Style Rules, Creating Styles for Nested Tags, Creating Classes and IDs for Applying Styles, Applying Styles to Hyperlinks, Creating and Linking to External Style Sheets **Formatting Text by Using Style Sheets:** Specifying a Font Family, Specifying a Font Size and Color, Applying Bold and Italics, Applying Strikethrough and Underlining, Creating Inline Spans , Adjusting Spacing Between Letters **Formatting Paragraphs by Using Style Sheets:** Indenting Paragraphs, Applying a Border to a Paragraph, Specifying a Border Style, Setting Border Padding, Specifying Border Width and Color, Formatting Border Sides Individually, Setting All Border Attributes at Once, Specifying the Horizontal Alignment of a Paragraph, Specifying Vertical Space within a Paragraph **Displaying Graphics:** Selecting a Graphics Format, Preparing Graphics for Web Use, Inserting Graphics, Arranging Elements on the Page, Controlling Image Size and Padding, Hyperlinking from Graphics, Using Thumbnail Graphics, Including Alternate Text for Graphics, Adding Figure Captions.	12
IV	**Java Script:** Introduction, Client-Side JavaScript, Server-Side JavaScript, JavaScript Objects, JavaScript Security Core JavaScript (Properties and Methods of Each) : Array, Boolean, Date, Function, Math, Number, Object, String, regExp **Document and its associated objects:** document, document object methods, Link, Area, Anchor, Image, Layer **Events and Event Handlers** : General Information about Events, Defining Event Handlers, event, onAbort, onBlur, onChange, onClick, onDblClick, onDragDrop, onError, onFocus, onKeyDown, onKeyPress, onKeyUp, onLoad, onMouseDown, onMouseMove, onMouseOut, onMouseOver, onMouseUp, onMove, onReset, onResize, onSelect, onSubmit, onUnload	12

V	**JSON**: Introduction, JSON Grammar, JSON Values, JSON Tokens, Syntax, JSON vs XML, Data Types, Objects, Arrays, Creating JSON, JSON Object, Parsing JSON, Persisting JSON, Data Interchange, JSON HTML, JSONP	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	HTML5 Step by Step	Faithe Wempen	Microsoft Press	-	2011
2	Web Design The Complete Reference	Thomas Powell	TMH	-	2009
3	Head First HTML5 programming	Eric Freeman	O'Reilly	-	2013
4	JavaScript 2.0: The Complete Reference	Thomas Powelland Fritz Schneider	TMH	2nd	2004
5	Beginning JSON	Ben Smith A	Apress	1st	2015

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Minor Project / Presentation

Web Technology Practical (SIUSDSP31)

List of Practical:

1.	**Use of Basic Tags:**
a.	Design a web page using different text formatting tags.
b.	Demonstrate use of Font tag with its attributes and HTML various color options in web page.
c.	Design a web page with links to different pages and allow navigation between web pages.
2.	**Navigation, list and paragraph:**
a.	Design a web page to demonstrate text-based navigation bar.
b.	Demonstrate use of lists and backgrounds in web page.
c.	Demonstrate use of paragraph and its associated tags in web page.
3.	**Lists, images and semantics:**
a.	Demonstrate use of multiple image tag in web page.
b.	Design a web page with Imagemaps.
c.	Design a web page demonstrating use of various semantics tags
4	**Multimedia and User controls:**
a.	Design a web page with a form that uses all types of user controls.
b.	Design a web page embedding with multimedia features.
c.	Design a 3 page static website with appropriate tags and attributes.
5	**CSS with list, links and table:**
a.	Create and use different style rules with available types of lists.
b.	Create and use different style rules with hyperlinks.
c.	Create and use different style rules with tables.
6	**CSS with font, paragraph and types:**
a.	Create and use different style rules with font elements.
b.	Create and use different style rules with Paragraph elements.
c.	Demonstrate the use of inline, internal and external CSS in one web page.
7	**Java Script: Validating User fields**
a.	Demonstrate the use of Document object methods

b.	Using java script, demonstrate validating Text Input Fields, Drop-down Lists and Checkboxes.
c.	Using java script, demonstrate validating Radio buttons and Validating Multi-Select Boxes.
8	**Java Script : Handling the events**
a.	Using java script, demonstrate the use of onAbort, onBlur, onChange, onClick, onDblClick events.
b.	Using java script, demonstrate the use of onDragDrop, onError, onFocus events.
c.	Using java script, demonstrate the use of onKeyDown, onKeyPress, onKeyUp, onLoad, onReset, onResize, onSelect, onSubmit, onUnload events
d.	Using java script, demonstrate the use of onMouseDown, onMouseMove, onMouseOut, onMouseOver, onMouseUp, onMove events.
e.	Using java script, demonstrate the use of onKeyDown, onKeyPress, onKeyUp, onLoad, onReset, onResize, onSelect, onSubmit, onUnload events.
f.	Demonstrate complete validation of User Registration form using appropriate fields of html and events of java script.
9	**JSON Basics**
a.	Creating JSON
b.	Parsing JSON
c.	Persisting JSON
10	**Working with JSON**
a.	Demonstrate use of JSON objects in array, print array on web page using document object.
b.	Read data from json file and convert it into a JavaScript object and display the data in web page using document object.
c.	Demonstrate messages formatting using JSON.

Data Structures and Program design Using Python (SIUSDS32)

Course Objective:

- To learn the essential Python data structures.
- To learn the significant Python implementation of popular data structures
- To learn about various data structure algorithms and design paradigms
- To acquire knowledge of how to create complex data structures.
- To acquire basic understanding of complex data structures such as trees and graphs and their applications

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Learner is capable of choosing appropriate data structure in Python for specified problems and algorithms.
CO2: Learner is able to implement Linked list and Stack data structure in various domains.
CO3: Learner is able to implement Tree and Queue data structures and use their operation.
CO4: Learner has ability to apply of Hashing techniques, Symbol Table and Graph Algorithms appropriately.
CO5: Learner has skills to handle sorting, searching and pattern matching on various data structures.

Unit	Contents	No. of Lectures
I	**Python Objects & Object-Oriented Programming:** Goals, Principles, and Patterns, Overview of data types and objects, Classes and object programming, Class Definitions, Inheritance, Data encapsulation and properties, Namespaces and Object-Orientation, Shallow and Deep Copying **Python Data Types and Structures:** Modules for data structures and algorithms-Collections, Deques, Chain Map objects Counter, Counter objects, Ordered dictionaries default dictionaries, Learning about named tuples Arrays **Principles of Algorithm Design:** An introduction to algorithms, Algorithm design paradigms Recursion and backtracking, Backtracking, Divide and conquer - long multiplication The recursive approach Runtime analysis Asymptotic analysis Big O notation, Composing complexity classes Omega notation, Theta notation, Amortized analysis	12
II	**Lists and Pointer Structures:** Arrays-Pointer structures **Singly linked lists**-Singly linked list class, The append operation, A faster append operation, Getting the size of the list, Improving list traversal, Deleting nodes, List search, Clearing a list **Doubly linked lists**-A doubly linked list node Doubly linked list class Append operation The delete operation List search **Circular lists**-Appending elements, Deleting an element in a circular list, Iterating through a circular list	12

	Stacks: Stack implementation, Push operation, Pop operation, Peek operation, Bracket-matching application	
III	**Queues:** List-based queues, Stack-based queues Node-based queues, Application of queues Media player queues **Trees:** Terminology, Tree nodes **Tree traversal:** Depth-first traversal In-order traversal and infix notation, Pre-order traversal and prefix notation, Post-order traversal and postfix notation, Breadth-first traversal **Binary search trees:** Binary search tree implementation, Binary search tree operations, Finding the minimum and maximum nodes Inserting nodes Deleting nodes, Searching the tree, Benefits of a binary search tree, Balancing trees, Expression trees, Parsing a reverse Polish expression **Heaps:** Ternary search tree	12
IV	**Hashing and Symbol Tables:** Hashing- Perfect hashing functions Hash tables-Storing elements in a hash table, Retrieving elements from the hash table, Testing the hash table, Using [] with the hash table, Non string keys, Growing a hash table, Open addressing, Chaining, Symbol tables **Graphs and Other Algorithms:** Graphs-Directed and undirected graphs, Weighted graphs, Graph representations, Adjacency lists, Adjacency matrices, Graph traversals- Breadth-first traversal, Depth-first search, Other useful graph methods, Priority queues and heaps- Insert operation, Pop Operation, Selection Algorithm	12
V	**Sorting:** Sorting algorithms- Bubble sort algorithms, Insertion sort algorithms, Selection sort algorithms, Quick sort algorithms **Selection Algorithms:** Selection by sorting, Randomized selection Quick Select, Deterministic selection-Pivot selection Median of medians Partitioning step **Pattern Matching Algorithms:** The brute-force algorithm, The Rabin Karp algorithm	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Data Structure using C	Reema Thareja	Oxford University Press	2nd	2014
2	Data Structure – A Pseudo Approach with C	Richard F. Gillberg	Cengage Publication	-	2005
3	Data Structures	Seymour Lipschutz	Tata McGraw Hill	-	2006

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Assignment / Presentation

Data Structures and Program Design Using Python Practical (SIUSDSP32)

List of Practical:

1.	**General Python Programs**
a.	Write Python Program to demonstrate the use of various Python Data Types and Structures
b.	Write Python Program to demonstrate OOP Concepts including Class, Objects, Inheritance and encapsulation.
c.	Write Python Program to implement array and operations of arrays.
2.	**List and Pointer Structure**
a.	Write Python Program to create singly linked list and various operations on it.
b.	Write Python Program to create doubly linked list and various operations on it.
c.	Write Python Program to create circular linked list and various operations on it.
3.	**Stacks and Queues**
a.	Write Python Program to implement stack and demonstrate push, pop and peek operations.
b.	Write Python Program to implement stack for Bracket-matching application
c.	Write Python Program to implement list based queues and demonstrate various operations on it.
d.	Write Python Program to implement stack based queues and demonstrate various operations on it.
e.	Write Python Program to implement Node based queues and demonstrate various operations on it.
f.	Write Python Program to implement queue data structure for simulating media player playlist queues.
4	**Trees**
a.	Write Python Program to implement tree data structure and demonstrate depth- first traversal
b.	Write Python Program to implement tree data structure and demonstrate breadth first traversal
c.	Write Python Program to implement binary search tree to find the minimum node
d.	Write Python Program to implement binary search tree to find the minimum node.
e.	Write a Python implementation to demonstrate the insert and delete method to add/delete the nodes in the BST.
f.	Python implementation to search the node in the BST

g.	Write a python program build up a tree for an expression written in postfix notation and evaluate it.
5	**Hashing and Symbol Tables**
a.	Write a Python Program to demonstration of computing Hash for given strings.
b.	Write a Python program to implement hash table for storing and searching values from it.
c.	Write aa Python Program to create Symbol Table
6	**Graphs**
a.	Write a Python program to store and display Graph data structure using adjacency matrix.
b.	Write a Python Program to implement Graph traversal (BFS/DFS) based on above practical.
c.	Write a Python program to implement priority queue and heap operations
7	**Searching**
a.	Write a Python Program for implementation in Python for the linear search on an unordered list of items
b.	Write a Python Program for implementation in Python for the linear search on an ordered list of items
c.	Write a Python Program for implementation of the binary search algorithm on an ordered list of items
d.	Write a Python Program for implementation of implementation of the interpolation search algorithm
8	**Sorting**
a.	Write a Python Program for implementing Insertion Sort.
b.	Write a Python Program for implementing Bubble Sort.
c.	Write a Python Program for implementing Quick Sort.
d.	Write a Python Program for implementing Selection Sort.
9	**Selection Algorithms**
a.	Write a Python Program to implement Randomized Selection
b.	Write a Python Program to implement Deterministic Selection
10	**Application**
a.	Write a Python Program to create an application for storing Polynomial
b.	Write a Python Program to create an application for adding two Polynomials

Probability and Distribution (SIUSDS33)

Course Objective:

- To explore about random variables and implement various distribution functions
- To familiarize with concepts of probability and learn implementation of different types of probabilities
- Learn and implement the concept of expectation, related theorems and generating functions
- To know the concept and implementation of discrete distributions including Bernoulli ,Binomial and power series distributions
- To get acquainted with theory and practical implementation of concepts of continuous distributions

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Organize, manage and present data
CO2: Analyse statistical data graphically using frequency distributions and cumulative frequency distributions
CO3: Use the basic probability rules, including additive and multiplicative laws, using the terms, in dependent and mutually exclusive events
CO4: Translate real-world problems into probability models
CO5: Calculate probabilities and derive the marginal and conditional distributions of bivariater and on variables.

Unit	Contents	No. of Lectures
I	**Theory of Probability:** Introduction, history, different terms, mathematical tools, Axiomatic approach to probability, Mathematical notation, multiplication and conditional probability, Bayes theorem, Geometric probability.	12
II	**Random Variables and Distribution Functions**: Random Variable, distribution function, discrete random variable, continuous rando m variable, joint probability law, transformation of one dimensional random variable, transformation of two dimensional random variable	12
III	**Mathematical Expectation and Generating Functions:** Mathematical expectation, Expectation of a Function of a Random Variable, Addition Theorem of Expectation, Multiplication Theorem of Expectation, Expectation of a Linear Combination of Random Variables, Covariance, Variance of a Linear Combination of Random Variables, Moments of Bivariate Probability Distributions, Conditional Expectation and Conditional Variance, Moment Generating Function, Cumulants, Characteristic Function, Cheby chev's Inequality, Convergence in- Probability, Weak Law of Large Numbers, Borel Canteli Lemma, Probability Generating	12

	Function	
IV	**Theoretical Discrete Distributions:** Introduction, Bernoulli Distribution ,Binomial Distribution, Poisson Distribution, Negative Binomial Distribution, Geometric Distribution, Hyper geometric Distribution, Multinomial Distribution, Discrete Uniform Distribution, Power Series Distribution	12
V	**Theoretical Continuous Distributions:** Rectangular or Uniform Distribution, Normal Distribution, Gamma Distribution, Beta Distribution of First Kind, Beta Distribution of Second Kind, The Exponential Distribution, Laplace Double Exponential Distribution, 12 28 Weibul Distribution, Cauchy Distribution, Central Limit Theorem, Compound Distributions ,Pearson\s Distributions, Variate Transformations, Order Statistics, Truncated Distribution.	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Fundamentals of Mathematical statistics	S.C, Gupta and V. K . Kapoor	S.Chand and Sons	Tenth	2002
2	Applied Statistics and Probability for Engineers	Douglas C. Montgomery and George C.Runger	Wiley	Sixth	2014
3	Probability, Statistics, and Stochastic Processes	Peter Olofsson and Mikael Andersson	Wiley	Second	2012

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Assignment / Tutorial Work

Probability and Distribution Practical (SIUSDSP33)

List of Practical : (Can be done in MS-Excel-or any Spreadsheet)

1.	**Introduction to Probability:**
a.	Formulate and apply Bayes' Theorem Calculations for problems like The "Two Supplier Example". [Hint: Use Prior Probabilities and Conditional Probabilities to compute Joint and Posterior Probabilities.]
b.	Design spreadsheet to demonstrate the association Between Two Variables by Computing the Covariance and Correlation Coefficient. [Hint: Use COVAR and CORREL)]
2.	**Discover Probability using formulas:**
a.	Design and spreadsheet experiment to compute the probability using the geometric distribution formula.
b.	Create a spreadsheet application to compute the Conditional Probability. Also determine the probability that a randomly chosen event.
3.	**Random Variables and Distribution Functions:**
a.	Create spreadsheet application to Compute the Expected Value, Variance, and Standard Deviation
b.	Create a spreadsheet application to Compute Binomial Probabilities. [Hint Use BINOMDIST]
4	**Probability Distribution and Law:**
a.	Create a spreadsheet application to Poisson Probability Distribution. [Hint: Use POISSON]
b.	Create a spreadsheet application to implement joint probability law.
5	**Mathematical Expectation and Cheby chev's Theorem:**
a.	Create a spreadsheet application to compute the expectation of a Function of a Random Variable
b.	Create a spreadsheet application to apply Cheby chev's Theorem.
6	**Conditional Expectation and Generating Functions:**
a.	Create a spreadsheet application to compute Conditional Expectation and Conditional Variance
b.	Create a spreadsheet application to demonstrate the use of Generating Functions
7	**Theoretical Discrete Distributions 1:**
a.	Create spreadsheet application to demonstrate Bernoulli Distribution.
b.	Create spreadsheet application to use excel function for computing hypergeometric probabilities.
8	**Theoretical Discrete Distributions 2:**

a.	Create spreadsheet application to Calculate Binomial Distribution in Excel. [Hint: Use BINOM.DIST]
b.	Create suitable spreadsheet application to work with Power Series Distribution.
9	**Theoretical Continuous Distributions 1:**
a.	Create spread application for computing probabilities and z values for the standard normal distribution. [Use NORMSDIST and NORMSINV]
b.	Create spread application for computing probabilities for the exponential probability distribution. [Hint: Use EXPONDIST]
10	**Theoretical Continuous Distributions 2:**
a.	Create spread application for demonstrating Weibull Distribution to obtain a model for data sets. [Hint: Use WEIBULL.DIST]
b.	Create spread application for demonstrating Pearson\s Distributions.

Testing of Hypothesis (SIUSDS34)

Course Objective:

- To impart statistical significance in solving complex problems.
- To critically test in developing robust, extensible and highly maintainable solutions to simple and complex problems.
- To implement various statistical functions using suitable programming languages and packages.
- To scientifically test the unknown and unlock possibilities in different dimensions of the system.
- To write the reports of analytical results generated by the system.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Learner is developing null and alternative hypotheses to test for a given situation.
CO2: Learner is able to differentiate one- and two-tailed hypothesis tests.
CO3: Learner is able to do sampling a normal distribution and random sampling.
CO4: Learner is using statistical models and their associations in performing hypothesis testing.
CO5: Lerner is writing the reports and interpreting the data using the various programming languages and packages.

Unit	Contents	No. of Lectures
I	**Introduction to Hypothesis Testing:** Hypothesis Tests, Stating a Hypothesis, Types of Errors and Level of Significance, Statistical Tests and P-Values, Making a Decision and Interpreting the Decision, Strategies for Hypothesis Testing, Characteristics of a good hypothesis, Steps for hypothesis testing **Hypothesis Testing for the Mean (σ Known):** Using P-Values to Make Decisions, Using P-Values for a z-Test, Rejection Regions and Critical Values, Using Rejection Regions for a z-Test, Critical Values in at Distribution, The t-Test for a Mean μ, Using P-Values with t-Tests, Sums and case studies **Packages used for Hypothesis testing:** Introduction to statistical functions in R / Python / Excel, Packages used for finding P-value to make decision and hypothesis testing.	12
II	**Goodness of fit tests:** Anderson-Darling, Chi-square test, Kolmogorov Smirnov, Ryan-Joiner, Shapiro-Wilk ,Jarque-Bera, Lilliefors **Variance tests:** Chi-square test of a single variance, F-tests of two variances, Tests of homogeneity **Wilcoxon rank-sum/Mann-Whitney U test, Sign test Contingency tables:** Chi-square contingency table test, G contingency table test, Fisher's exact test, Measures of association, McNemar's test **Packages used for Hypothesis testing:** Packages used for finding goodness of fit test, variance test, Wilcoxon rank-sum / Mann-Whitney U test and sign test, Using Contingency table in R / Python / Excel.	12
III	**Analysis of variance and covariance:** ANOVA, Single factor or one way ANOVA, Two factor or two-way and higher-way ANOVA , MANOVA, ANCOVA **Non-Parametric ANOVA:** Kruskal-Wallis ANOVA, Friedman ANOVA test, Mood's Median	12

	Packages used for Hypothesis testing: Packages used for finding Anova, Manova, Anova and Non-Parametric Anova in R / Python / Excel.	
IV	**Regression and smoothing:** Least squares, Ridge regression, Simple and multiple linear regression, Polynomial regression, Generalized Linear Models (GLIM), Logistic regression for proportion data, Poisson regression for count data, Non-linear regression, Smoothing and Generalized Additive Models (GAM), Geographically weighted regression (GWR), Spatial series and spatial autoregression- SAR models, CAR models, Spatial filtering models **Time series analysis and temporal auto regression:** Moving averages, Trend Analysis, ARMA and ARIMA (Box-Jenkins) models, Spectral analysis	12
V	**Communicating and Documenting the Results of Analyses:** Introduction, The Difficulty of Good Communication, Communication Hurdles: Graphical Distortions, Communication Hurdles: Biased Samples & Sample Size, Preparing Data for Statistical Analysis, Guidelines for a Statistical Analysis and Report, Documentation and Storage of Results ,Supplementary Exercise **Data Storytelling:** What is a Data Story?, The Art and Science of Storytelling, Planning the Data Story, Elements of the Data Story, Parts of the Data Story, Framing and Formatting of the Data Story, False Narratives and Data Storytelling **Infographics:** What is an Infographic?, Why are Infographics Useful? Types of Infographics, Infographic Design Elements, Steps in Designing an Infographic, Best Practices in Designing an Infographic	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Hypothesis Testing	---	Pearson Higher Education	---	---
2	Statistical Analysis Handbook	Dr. Michael J de Smith	The Winchelsea Press, Drumlin Security Ltd, Edinburgh	---	2018
3	An Introduction to Statistical Methods and Data Analysis	R. Lyman Ott & Michael Longnecker	Thomson Learning	---	---

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Assignment / Tutorial Work

Testing of Hypothesis Practical (SIUSDSP34)

List of Practical : (Practical can be performed using R / Python / scilab / matlab / SPSS / MS Excel)

1	**Hypothesis Testing for the Mean**
a.	Perform testing of hypothesis using one sample t-test.
b.	Perform testing of hypothesis using two sample t-test.
c.	Perform testing of hypothesis using paired t-test.
d.	Perform testing of hypothesis using Z-test.
2	**Goodness-of-fit test**
a.	Perform goodness-of-fit test using chi-squared test.
b.	Perform goodness-of-fit test using KS-test.
c.	Perform testing of hypothesis using chi-squared Test of Independence
3	**Variance Testing**
a.	Using Chi-square test of a single variance
b.	Using F-tests of two variances
c.	Testing of homogeneity
4	**Analysis of variance and covariance**
a.	Perform testing of hypothesis using one-way ANOVA.
b.	Perform testing of hypothesis using two-way ANOVA.
c.	Perform testing of hypothesis using Multivariate ANOVA (MANOVA)
d.	Perform testing of hypothesis using one-way ANOVA.
5	**Regression**
a.	Perform simple linear regression
b.	Perform multiple linear regression
c.	Perform polynomial regression
6	Perform spatial series and spatial auto-regression
7	Perform time series analysis using Moving averages
8	Perform time series analysis using Trend Analysis
9	Perform Spectral analysis
10	Creating Infographics‖ using secondary data available on internet. (Use Canva / Adobe Spark / Prezi / Vennage

Introduction to Business Analytics (SIUSDS35)

Course Objectives:

- To develop core abilities to make data-driven decisions through big data.
- To provide an overview of an exciting growing field of big data analytics.
- To introduce the tools required to manage and analyze big data like Hadoop, NoSql MapReduce.
- To teach the fundamental techniques and principles in achieving big data analytics with scalability and streaming capability.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Learner understands the key issues in big data management and its associated applications in intelligent business and scientific computing.
CO2: Lerner is acquiring fundamental techniques and algorithms like Hadoop, Map Reduce and NO SQL in big data analytics.
CO3: Learner is able to interpret business models and scientific computing paradigms, and apply software tools for big data analytics.
CO4: Learner understands adequate perspectives of big data analytics in various applications like recommender systems, social media applications etc.

Unit	Contents	No. of Lectures
I	**Introduction to Big Data Analytics:** Defining Big Data analytics: Discovering value from large data sets, Exploiting data to optimize decision-making **Planning your analytics life cycle project:** Outlining steps in the life cycle, Contrasting traditional analytics with Big Data analytics Representing Big Data with R and Rattle: **Preparing the data:** Loading data for knowledge discovery, Spotting outliers in the data, Transforming and summarizing data Visualizing data characteristics: Revealing changes over time, Displaying proportions within your data, Leveraging charts to display relationships, Displaying relationships across categories	12
II	**Modeling and Predictive Data Analysis:** Categorizing analytic approaches: Predictive vs. descriptive analytics, Supervised vs. unsupervised learning **Applying appropriate mining techniques:** Discovering unknown groups through clustering, Detecting relationships with association rules, Uncovering decision tree classifications, Identifying patterns with time series analysis	12
III	**Leveraging Analytics with RHadoop** Expanding the analytic capabilities of your organization Exploring the Map Reduce and Hadoop architecture, Creating and executing Hadoop Map Reduce jobs, Integrating R and Hadoop with RHadoop, Examining	12

	the components of RHadoop, Creating modules for RHadoop jobs, Executing RHadoop jobs, Monitoring job execution flow **Building a Recommendation Framework:** Streamlining business decisions Considering motivations for a recommender engine, Leveraging recommendations based on collaborative filtering, Exploring the architecture of the recommendation framework, Building programming components, Executing the recommendation model, Performing tradeoff analysis	
IV	**Mining Unstructured Data** Investigating business value within unstructured data Making a business case for unstructured data mining, Extending business intelligence with mining tools Implementing text mining and social network analysis Analyzing the structure of text mining, Evaluating mining approaches, Building a text mining framework, Inspecting social network interactions	12
V	**Planning and Implementing a Complete Data Analytics Solution** Transforming business objectives to analytic projects Arguing your business case for analytics, Mapping analytics models to business objectives, Identifying performance metrics targets Implementing the analytics life cycle Finding core data sets, Preparing the data for analysis, Modeling the data, Executing the model, Communicating results Ensuring a Successful Data Analytics Solution Identifying barriers to Big Data analytics, Managing and mitigating risks, Employing an implementation checklist	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Big Data Analytics with R and Hadoop	VigneshPrajapati	Packt Publishing House	---	---
2	Analytics in a Big Data World: The Essential Guide to Data Science and its Applications	Bart Baesens	Wiley	---	---
3	Big Data Analytics Disruptive Technologies For Changing The Game	ArvindSathi	MC Press LLC	---	---

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Assignment / Tutorial Work

Introduction to Business Analytics (SIUSDSP35)

List of Practical:

1	**Installation Cloud Databases**
a.	Install, configure and run Hadoop and HDFS
b.	Implement word count/ frequency program using MapReduce
2	Implement an Map reduce program that process a weather dataset
3	Exploring Hadoop Distributed File System (HDFS)
4	Implement an application that store big data in Hbase/ Mongodb/ Pig using Hadoop/R
5	Implement a program in Pig
6	Configure the Hive and implement the application in Hive
7	Illustrate the working of Jaql
8	**Implement Classification Technique**
a.	Implement Decision tree classification technique
b.	Implement SVM Classification technique
9	**Implement Regression Model**
a.	Regression Model: Import a data from web storage. Name the dataset and do Logistic Regression to find out relation between variables that are affecting the admission of a student in an institute based on his or her GRE score, GPA obtained and rank of the student. Also check the model is fit or not require (foreign), require (Mass)
b.	MULTIPLE REGRESSION MODEL: Apply multiple regressions, if data have a continuous independent variable. Apply on above dataset.
10	**Implement Classification and Clustering Model**
a.	CLASSIFICATION MODEL: a. Install relevant package for classification. b. Choose classifier for classification problem. c. Evaluate the performance of classifier.
b.	CLUSTERING MODEL a. Clustering algorithms for unsupervised classification. b. Plot the cluster data using R visualizations.

SEMESTER - 4

SYBSc(DS) - Semester IV

Course Code	Course Type	Course Title	Periods per week (1 Period = 50 min)	Credits
SIUSDS41	Core Subject	Data warehousing and Mining	5	2
SIUSDS42	Core Subject	Linear Algebra	5	2
SIUSDS43	Skill Enhancement Course	R Programming	5	2
SIUSDS44	Core Subject	Research Methodology	5	2
SIUSDS45	Core Subject	Optimization Techniques	5	2
	Core Subject			
SIUSDSP41	Core Subject Practical	Data warehousing and Data Mining Practical	3	2
SIUSDSP42	Core Subject Practical	Linear Algebra Practical	3	2
SIUSDSP43	Skill Enhancement Course Practical	R Programming Practical	3	2
SIUSDSP44	Core Subject Practical	Research Methodology Practical	3	2
SIUSDSP45	Core Subject Practical	Optimization Techniques Practical	3	2
		TOTAL CREDITS		20

Data warehousing and Data Mining (SIUSDS41)

Course Objective:

- Understand business intelligence for an enterprise and review data warehouse with architectural types and architectural building blocks
- To discuss and understand changing dimensions and learn about aggregate tables and determine their usage.
- To learn basics of data mining, understand the need and the process of data mining in contrast with machine learning.
- To study the use of classification and clustering techniques for Data Mining.
- To appreciate the use of various data mining algorithms and learn about their specific applications.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: :Learner is able to demonstrate knowledge of business intelligence, data warehouse with clear understanding of architectural types and will be able to establish the relationship between architectural building blocks.
CO2: Learner is able to elaborate changing dimensions with respect to current trends & using aggregate tables.
CO3: Learner is able to handle the processes of data preprocessing, data transformation and data reduction.
CO4: Learner has knowledge of using various Data Mining techniques for classification and clustering.
CO5: Learner is able to align the Data Mining techniques for analyzing the datasets using tools like Weka, R or Python.

Unit	Contents	No. of Lectures
I	**THE COMPELLING NEED FOR DATA WAREHOUSING:** Escalating Need for Strategic Information, Failures of Past Decision-Support Systems, Operational Versus Decision-Support Systems, Data Warehousing—The Only Viable Solution, Data Warehouse Defined, The Data Warehousing Movement, Evolution of Business Intelligence **DATA WAREHOUSE:** The Building Blocks: Defining Features, Data Warehouses and Data Marts, Architectural Types, Overview of The Components, Metadata in The Data Warehouse **TRENDS IN DATA WAREHOUSING:** Continued Growth in Data Warehousing, Significant Trends, Emergence of Standards, Web-Enabled Data Warehouse **ARCHITECTURAL COMPONENTS:** Understanding Data Warehouse Architecture, Distinguishing Characteristics, Architectural Framework, Technical Architecture, Architectural Types **THE SIGNIFICANT ROLE OF METADATA:** Why Metadata Is Important, Metadata Types By Functional Areas, Business Metadata, Technical Metadata, How To Provide Metadata	12

II	**PRINCIPLES OF DIMENSIONAL MODELING:** From Requirements to Data Design, The Star Schema, Star Schema Keys, Advantages of The Star Schema, Star Schema: Examples **DIMENSIONAL MODELING: ADVANCED TOPICS:** Updates to The Dimension Tables, Miscellaneous Dimensions, The Snowflake Schema, Aggregate Fact Tables, Families of Stars **DATA EXTRACTION, TRANSFORMATION, AND LOADING:** ETL Overview, ETL Requirements and Steps, Data Extraction, Data Transformation, Data Loading, ETL Summary, Other Integration Approaches	12
III	**INTRODUCTION TO DATA MINING:** Introduction to Data Mining, Need of Data Mining, What Can Data Mining Do and Not Do? Data Mining Applications, Data Mining Process, Data Mining Techniques, Difference between Data Mining and Machine Learning **BEGINNING WITH WEKA AND IRIS DATASET IN R:** About Weka, Installing Weka, Understanding Fisher's Iris Flower Dataset, Preparing the Dataset, Understanding A RFF, Working with a Dataset in Weka, Working with the Iris dataset in R Data Preprocessing: Need for Data Preprocessing, Data Preprocessing Methods **CLASSIFICATION:** Introduction to Classification, Types of Classification, Input and Output Attributes, Guidelines for Size and Quality of the Training Dataset, Introduction to the Decision Tree Classifier, Naive Bayes Method, Understanding Metrics to Assess the Quality of Classifiers	12
IV	**CLUSTER ANALYSIS:** Introduction to Cluster Analysis, Applications of Cluster Analysis, Desired Features of Clustering, Distance Metrics, Major Clustering Methods/Algorithms, Partitioning Clustering, **HIERARCHICAL CLUSTERING ALGORITHMS:** Web Mining and Search Engines: Introduction, Web Content Mining, Web Usage Mining, Web Structure Mining, Hyperlink Induced Topic Search algorithm, Introduction to Modern Search Engines, Working of a Search Engine, PageRank Algorithm, Precision and Recall	12
V	**INTRODUCTION TO ASSOCIATION RULE MINING:** Defining Association Rule Mining, Representations of Items for Association Mining, The Metrics to Evaluate the Strength of Association Rules, The Naive Algorithm for Finding Association Rules, Approaches for Transaction Database Storage **THE APRIORI ALGORITHM,** Closed and Maximal Item sets, The Apriori– TID Algorithm for Generating Association Mining Rules, Direct Hashing and Pruning (DHP), Dynamic Itemset Counting (DIC), Mining Frequent Patterns without Candidate Generation (FP Growth)	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	DATA WAREHOUSING FUNDAMENTALS FOR IT PROFESSIONALS	PAULRAJ PONNIAH	Wiley	Second	2010
2	Data Mining and Data Warehousing : Principles and Practical Techniques	Parteek Bhatia	Cambridge University Press	First	2019
3	Data Mining and Data Warehousing	S.K. Mourya Shalu Gupta	Alpha Science International Ltd	First	2013

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Assignment / Presentation

Data warehousing and Data Mining Practical (SIUSDSP41)

List of Practical:

1.	**Data warehouse design**
a.	Design dimension tables.
b.	Design fact tables.
c.	Create an indexed view and rebuild column store indexes.
2.	**Data Warehouse with Azure**
a.	Create an Azure SQL Data Warehouse Project.
b.	Develop tables in Azure SQL Data Warehouse.
c.	Migrate Data Warehouse to Azure.
d.	Pause and remove Azure data warehouse.
3.	**Data Warehouse implementation and use**
a.	Cleanse data with SQL Server Data Quality Services.
b.	Create custom knowledge base.
c.	Install Master Data Services and IIS
d.	Configure MDS and deploy sample MDS model.
e.	Install MDS excel add-in and Update master data in excel.
f.	Consume the data from the warehouse.
4	**Working with Data and Data Preprocessing**
a.	Demonstrate the use of ARFF files taking input and display the output of the files.
b.	Create your own excel file. Convert the excel file to .csv format and prepare it as ARFF files.
c.	Preprocess and classify Customer dataset.http://archive.ics.uci.edu/ml/
d.	Perform Preprocessing, Classification techniques onAgriculture dataset. (http://archive.ics.uci.edu/ml/)
e.	Preprocess and classify Weather dataset.http://archive.ics.uci.edu/ml/
f.	Perform data Cleansing of customer dataset. http://archive.ics.uci.edu/ml/ www.kdnuggets.com/datasets/

5	**Performing classification on data sets**
a.	Building a Decision Tree Classifier in Weka
b.	Applying Naïve Bayes on Dataset for classification
c.	Creating the Testing Dataset
d.	Decision Tree Operation with R
e.	Naïve Bayes Operation using R
f.	Classify the dataset using decision tree. www.kdnuggets.com/datasets/
6	**Simple Clustering**
a.	Perform Clustering technique on Customer dataset. http://archive.ics.uci.edu/ml/
b.	Perform Clustering technique on Agriculture dataset. http://archive.ics.uci.edu/ml/
c.	Perform Clustering technique on Weather dataset. http://archive.ics.uci.edu/ml/
7	**Implementing Clustering with Weka and R**
a.	Clustering Fisher‘s Iris Dataset with the Simple k-Means Algorithm
b.	Handling Missing Values
c.	Results Analysis after Applying Clustering
d.	Classification of Unlabeled Data
e.	Clustering in R using Simple k-Means
8	**Implementing Apriori Algorithm with Weka and R**
a.	Applying Predictive Apriori in Weka
b.	Applying the Apriori Algorithm in Weka on a Real World Dataset
c.	Applying the Apriori Algorithm in Weka on a Real World Larger Dataset
d.	Applying the Apriori Algorithm on a Numeric Dataset
9	**Implementing Association Mining with R**
a.	Applying Association Mining in R
b.	Application of Association Mining on Numeric Data in R

c.	Perform Association technique on Agriculture dataset. http://archive.ics.uci.edu/ml/,www.kdnuggets.com/datasets/
d.	Perform Association technique on Agriculture dataset. http://archive.ics.uci.edu/ml/, www.kdnuggets.com/datasets/
e.	Perform Association technique on Weather dataset.
10	**Web Mining**
a.	Implement Hyperlink Induced Topic Search (HITS) algorithm
b.	Implement PageRank Algorithm

Linear Algebra (SIUSDS42)

Course Objective:

- To analyze the solution set of a system of linear equations.
- To interpret existence and uniqueness of solutions geometrically.
- To formulate, solve, apply, and interpret properties of linear systems.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Learner is able to perform common matrix operations such as addition, scalar multiplication, multiplication, and transposition.
CO2: : Learner is able to describe how the determinant of a product of matrices relates to the determinant of the individual matrices.
CO3: : Learner expresses clear understanding of the concept of a _solution to a game' and also the limitations on the applicability of the theory

Unit	Contents	No. of Lectures
I	**Matrices and Gaussian Elimination:** Introduction, The Geometry of Linear Equations, An Example of Gaussian Elimination, Matrix Notation and Matrix Multiplication, Triangular Factors and Row Exchanges, Inverses and Transposes, Special Matrices and Applications **Vector Spaces:** Vector Spaces and Subspaces, Solving Ax= 0 and Ax= b, Linear Independence, Basis, and Dimension, The Four Fundamental Subspaces, Graphs and Networks, Linear Transformations	12
II	**Orthogonality:** Orthogonal Vectors and Subspaces, Cosines and Projections onto Lines, Projections and Least Squares, Orthogonal Bases and Gram-Schmidt, The Fast Fourier Transform **Determinants:** Introduction, Properties of the Determinant, Formulas for the Determinant, Applications of Determinants	12
III	**Eigenvalues and Eigenvectors:** Introduction, Diagonalization of a Matrix, Difference Equations and Powers Ak, Differential Equations and eAt, Complex Matrices, Similarity Transformations	12
IV	**Positive Definite Matrices:** Minima, Maxima, and Saddle Points, Tests for Positive Definiteness, Singular Value Decomposition, Minimum Principles, The Finite Element Method **Computations with Matrices:** Introduction, Matrix Norm and Condition Number, Computation of Eigenvalues, Iterative Methods for Ax= b	12

V	**Linear Programming and Game Theory:** Linear Inequalities, The Simplex Method, The Dual Problem, Network Models, Game Theory	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Linear Algebra and Its Applications	Gilbert Strang	Cengage Publication	Fourth Edition	---
2	Advanced Linear Algebra	David Surowski			
3	Linear Algebra, Theory and Applications	Kenneth Kuttlet			

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Assignment / Tutorial Work

Linear Algebra Practical (SIUSDSP42)

List of Practical:

1	**Matrices and Gaussian Elimination**.
a.	Multiplication and transpose of matrix using R/ python/ scilab/ matlab.
b.	Inverses of matrix in R/ python/ scilab/ matlab without using any inbuilt package.
c.	Inverses of matrix in R/ python/ scilab/ matlab using any inbuilt package like numpy.
d.	Linear equation with n unknowns using Gauss Elimination Method using R/python/ scilab/ matlab
2	**Vector**
a.	Addition, subtraction, multiplication and division of vector using R/ python/ scilab/ matlab.
b.	dot product & cross product of vector using R/ python/ scilab/ matlab
c.	Visualising vector Linear Transformations using R/ python/ scilab/ matlab
3	
a.	Computes the orthonormal vectors using the GS algorithm using R/ python/ scilab/ matlab.
b.	Projections and Least Squares using R/ python/ scilab/ matlab.
c.	Fast Fourier Transform using R/ python/ scilab/ matlab.
4	
a.	Finding determinant of matrix in R/ python/ scilab/ matlab without using any inbuilt package.
b.	Finding determinant of matrix in R/ python/ scilab/ matlab using any in built package.
5	
a.	Compute the eigenvalues and right eigenvectors of a given square array using R/ python/ scilab/ matlab.
b.	Program to test diagonalizable matrix using R/ python/ scilab/ matlab.
6	
a.	Tests for Positive Definiteness using R/ python/ scilab/ matlab.
b.	Singular Value Decomposition using R/ python/ scilab/ matlab.
c.	The Finite Element Method using R/ python/ scilab/ matlab. (**Only Demonstration**)
7	Simplex Method using R/ python/ scilab/ matlab. (**Only Demonstration**)
8	The Dual Problem using R python/ scilab/ matlab. (**Only Demonstration**)

9	Implementing Network Models using R/ python/ scilab/ matlab. (**Only Demonstration**)
10	Implementing Game Theory using R/ python/ scilab/ matlab. (**Only Demonstration**)

R Programming (SIUSDS43)

Course Objective:

- Master the use of the R interactive environment and expanding by installing R packages
- Read Structured Data into R from various sources
- Understand the different data types and data structures in R
- Manipulate strings, dates in R
- Understand basic regular expressions in R
- Understand base R graphics
- Focus on GGplot2 graphics for R and be familiar with trellis (lattice) graphics.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: To use R Studio and explore the features for R programming• To apply formatting on table, use Pipelines in application and use strings, factors in R programme.
CO2: To use R functions and graphics with in R programming for solving problems.
CO3: . To work with advanced graphics of R, import and use the data and represent the data into tables.
CO4: To manipulating Data Frames and make use of Dates in R application.

Unit	Contents	No. of Lectures
I	**Getting started with R:** R Software: Obtaining R and RStudio, First R Encounter, Getting started: R as a big calculator, Assignment, Basic operators, Help with functions and features, Quiz, A few important points on R Working with R **R Interfaces** - Using R and RStudio: R Software, Obtaining R and RStudio, The default R interface, RStudio Interface, Example Datasets in R, R Packages, Installing new R libraries, Customizing R Start-up **Objects in R:** Using ls and rm to managing R Objects, Types of R objects, Attributes of R Objects, Creating and accessing objects, Modifying elements, Quick recap, Exercise **Reading and writing data to and from R:** Importing and reading text files data into RStudio, Importing data using R command read.table(), Exercise, Importing text files Using scan(), Parsing each line – Readlines, Writing Data table from R, Exercise, Importaing Data from other Software, Reading data from Excel into R, Import/Export from other statistical software, From a Database Connection, Sampling and Creating simulated data, Exercise	12
II	**Introduction to programming and writing Functions in R:** Why do we want to write functions?, Conditional statements (if, ifelse, switch), Repetitive execution: For and While loops, The Apply Functions, Exercise, Functions for parsing text, Programming in R: More advanced, Viewing Code of functions from R packages, ExerciseParsing Real Data - World Population Data from Wikipedia, Writing functions: more technical discussion -Scoping, Options for Running memory or CPU intensive jobs in R, Efficient R coding	12

	Introduction to graphics in R: The R function plot(), Exercise, Customize plot with low-level plotting commands, Default parameters – par, Interacting with graphics, Saving plots, Useful Graphics Resources	
III	**Advanced Graphics:** Advanced plotting using Trellis; ggplots2, Lattice, Examples that Present Panels of Scatterplots using xyplot(), Simple use of xyplot **Importing Data- readr:** Functions for Reading Data, File Headers, Column Types, String-based Column Type Specification, Functionbased Column Type Specification Parsing Time and Dates, Space-separated Columns, Functions for Writing Data **Representing Tables – tibble**: Creating Tibbles, Indexing Tibbles	12
IV	**Reformatting Tables – tidyr:** Tidy Data, Gather and Spread, Complex Column Encodings, Expanding, Crossing, and Completing, Missing Values, Nesting **Data Pipelines – magrittr :** The Problem with Pipelines, Pipeline Notation, Pipelines and Function Arguments, Function Composition, Other Pipe Operations **Working with Strings – stringr:** Counting String Patterns, Splitting Strings, Capitalizing Strings, Wrapping, Padding, and Trimming, Detecting Substrings, Extracting Substrings, Transforming Strings **Working with Factors – forcats:** Creating Factors, Concatenation, Projection, Adding Levels, Reorder Levels	12
V	**Manipulating Data Frames – dplyr:** Selecting Columns, Filter, Sorting, Modifying Data Frames, Grouping and Summarizing, Joining Tables, Income in Fictional Countries **Working with Dates – lubridate**: Time Points, Time Zones, Time Intervals	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Introduction to Programming and Statistical Modelling in R	Aedin Culhane	HARVARD SCHOOL	1 st	2013
2	R Data Science Quick Reference	Thomas Mailund	Apress	1 st	2019
3	Beginning Data Science in R	Thomas Mailund	Apress	---	2017

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Assignment / Project

R Programming Practical (SIUSDSP43)

List of Practical:

1	**Introduction to R Programming Elements**
a.	Write an R Program to implement expressions, assignment and decision making
b.	Write an R Program to design and implement loops.
c.	Write a R program to demonstrate the use of essential data structures in R [Hint: Vectors, Matrix, Arrays]
2	**Using List, Data Frames and Functions in R**
a.	Write an R program to manage data and exhibit operations on it using List data structure
b.	Write an R program to manage data and exhibit operations on it using Data Frames
c.	Write an R program to demonstrate the use of : i. user-defined functions ii. built-in numeric function, character functions etc.
3	**Implementing Strings in R**
a.	Write an R program to store and access string in R objects(vectors, matrix, arrays, data frames, and lists)
b.	Write an R program to demonstrate use of various string manipulation functions. [Hint: paste(), print(), noquote(),format(), cat(), toString(), sprint()]
4	**Performing Statistics with R-I**
a.	Write an R program to apply built-in statistical functions. [Hint: mean, median, standard deviation and others]
b.	Write an R program to demonstrate Linear and Multiple Regression analysis.
5	**Performing Statistics with R-II**
a.	Write an R program to implement i. Normal Distribution. [Hint: dnorm(), pnorm(), qnorm(), rnorm()] ii. Binomial Distribution: [Hint: dbinom(), pbinom(),qbinom(),rbinom()]
b.	Write an R program to perform time-series analysis for the given data.
6	**Data Visualization and Analysis**
a.	Write an R program to learn about Tabulation and related concepts [Hint: Contingency Tables, Selection of Parts, Conversion, Complex Tables, Cross Tabulation]

b.	Write an R program to demonstrate various ways of performing Graphical analysis.[Hint: Plots, Special Plots, Storing Graphics
7	**Object Oriented Programming in R**
a.	Write an R program to demonstrate OOP concepts, the construction and use of S3 and S4 classes.
b.	Write an R program to define reference class and operations on them
8	**Data Interfaces in R**
a.	Write an R program to demonstrate data interface with CSV files [Hint: creating data for CSV, analyzing, writing CSV files]
b.	Write an R program to work with spreadsheet (Excel) programs. [Hint: installing, loading, verifying, creating data for xlsx file]
c.	Write an R program to manage data using XML files. [Develop data interface for maintaining Employee Information]
d.	Write an R program to demonstrate working with RMySQL Package
9	**Handling Errors in R**
a.	Write an R program to demonstrate various error messages in R Programming
b.	Write an R program to implement Error Handling in R [Hint: warning(),stop(),try(), tryCatch(), CallingHandlers()]
10	**Measuring Performance**
a.	Write R program to measure the performance with the help of bult-in function like mocrobechmark().

Research Methodology (SIUSDS44)

Course Objective:

- To impart analytical skill in solving complex problems.
- To foster the ability to critically think in developing robust, extensible and highly maintainable solutions to simple and complex problems.
- To explore the unknown and unlock new possibilities in different dimensions of the system.
- To portray accurately the characteristics of a particular individual, situation or a group under study

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Learner understands the reasons for doing research, the applications of research, characteristics and requirements of the research process, types of research and Research paradigms.
CO2: Learner is applying major approaches to information gathering, the relationship between attitudinal and measurement scales Methods for exploring attitudes in research.
CO3: Learner is able to analyze data in qualitative and quantitative studies, application of IT in data analysis.
CO4: Learner is able to write a research report and use Information Technology in Research
CO5: :Learneris practicing ethical codes and practices of conduct research.

Unit	Contents	No. of Lectures
I	**Research Methodology:** An Introduction: Meaning of Research, Objectives of Research, Motivation in Research, Types of Research, Research Approaches, Significance of Research, Research Methods versus Methodology, Research and Scientific Method, Importance of Knowing How Research is Done, Research Process, Criteria of Good Research, Problems Encountered by Researchers in India **Defining the Research Problem:** What is a Research Problem?, Selecting the Problem, Necessity of Defining the Problem, Technique Involved in Defining a Problem, An Illustration **Research Design:** Meaning of Research Design, Need for Research Design,Features of a Good Design, Important Concepts Relating to Research Design, Different Research Designs, Basic Principles of Experimental Designs.	12
II	**Sampling Design:** Census and Sample Survey, Implications of a Sample Design, Steps in Sampling Design, Criteria of Selecting a Sampling Procedure, Characteristics of a Good Sample Design, Different Types of Sample Designs, How to Select a Random Sample?, Random Sample from an Infinite Universe, Complex Random Sampling Designs **Measurement and Scaling Techniques:** Measurement in Research, Measurement Scales, Sources of Error in Measurement, Tests of Sound Measurement, Technique of Developing Measurement Tools, Scaling, Meaning of Scaling, Scale Classification Bases, Important Scaling Techniques, Scale Construction	12

	Techniques Methods of Data Collection: Collection of Primary Data, Observation Method, Interview Method, Collection of Data through Questionnaires, Collection of Data through Schedules, Difference between Questionnaires and Schedules, Some Other Methods of Data Collection, Collection of Secondary Data, Selection of Appropriate Method for Data Collection, Case Study Method, (i) Guidelines for Constructing Questionnaire/Schedule (ii) Guidelines for Successful Interviewing (iii) Difference between Survey and Experiment	
III	**Processing and Analysis of Data:** Processing Operations, Some Problems in Processing, Elements/Types of Analysis, Statistics in Research, Measures of Central Tendency, Measures of Dispersion, Measures of Asymmetry (Skewness), Measures of Relationship, Simple Regression Analysis, Multiple Correlation and Regression, Partial Correlation, Association in Case of Attributes, Other Measures, Summary Chart Concerning Analysis of Data **Sampling Fundamentals:** Need for Sampling, Some Fundamental Definitions, Important Sampling Distributions, Central Limit Theorem, Sampling Theory, Sandler's A-test, Concept of Standard Error, Estimation, Estimating the Population Mean (m), Estimating Population Proportion, Sample Size and its Determination, Determination of Sample Size through the Approach Based on Precision Rate and Confidence Level, Determination of Sample Size through the Approach, Based on Bayesian Statistics **Testing of Hypotheses:** What is a Hypothesis?Basic Concepts Concerning Testing of Hypotheses, Procedure for Hypothesis Testing, Flow Diagram for Hypothesis Testing, Measuring the Power of a Hypothesis Test, Tests of Hypotheses, Important Parametric Tests, Hypothesis Testing of Means, Hypothesis Testing for Differences between Means, Limitations of the Tests of Hypotheses	12
IV	**Interpretation of Data and Paper Writing –** Layout of a Research Paper, Journals in Computer Science, Impact factor of Journals, When and where to publish ?, UGC-CARE, Web of Science, SCOPUS, IEEE, ACM, Ethical issues related to publishing, Copyright, Data Privacy, Plagiarism and Self-Plagiarism, Software for detection of Plagiarism. ShodhShudhhi (PDS), smallseotools.com Use of Encyclopedias, Research Guides, Handbook etc., Academic Databases for Computer Science and Information Technology Discipline. Google Scholar, shodhganga, IEEE Xplore, ResearchGate, IDELS, DASH **Use of tools / techniques for Research:** Chicago, Turabian, MLA and APA Style, Reference Management Software like EndNote, Zotero or Mendeley; Software for paper formatting like LaTeX/MS Office/ **Scrivener/Open Office/Google Doc/DropBox Paper.**	12

V	**Ethics in business research:** What Are Research Ethics? Ethical Treatment of Participants, Ethics and the Sponsor, Researchers and Team Members, Professional Standards, Resources for Ethical Awareness **Think like a Researcher:** The Language of Research, Concepts, Constructs, Definitions, Variables, Propositions and Hypotheses, Theory, Models, Research and the Scientific Method, Sound Reasoning for Useful Answers **E-Research:** Introduction, The Internet as object of analysis, Using websites to collect data from individuals. Virtual ethnography, Qualitative research using online focus groups, Qualitative research using online personal interviews, Online social surveys, Ethical considerations in e-research, The state of e-research	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Research Methodology – Methods and techniques	C. R. Kothari	New Age International (P) Ltd., Publishers	---	---
2	Business Research Methods	Donald R. Cooper Pamela Schindler	Mc Graw Hill/ Irwin	12th Ed	---
3	RESEARCH METHODOLOGY - a step by step guide for beginners	Ranjit Kumar	SAGE Publication Ltd	---	---

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Research Paper Writing / Published

Research Methodology Practical (SIUSDSP44)

List of Practical:

1	**Introduction to LaTex**
a.	**Report Writing:** report style having chapter, section and subsection, article style having section, subsection and subsubsection, Automatic generation of table of contents, toc file to store the information that goes into the table of contents, Automatic numbering of section numbers.
b.	**Equations and Numbering Equations:** Creating an equation, writing multiple equations, Aligning multiple equations, creating matrices in Latex, label command, Cross referencing with ref command.
c.	**Tables and Figures:** Tables and Figures Creating tables and figures in LaTeX
d.	**Bibliography:** Bibliography Creating Bibliography in LaTeX
2	**Introduction to EndNote, Zotero or Mendeley**
a.	Integration with Word and adding citation and creating bibliographies
b.	Creating your own library
c.	Creating references from website
d.	Creating references manually
3	Visit the college library or nearby research center or from internet collect 5 tittles of research papers/thesis and classify them according to types of research, Discuss how the problems are delineated, how they are relevant to scientific method etc.
4	Identify 2 researchable problems relevant to your context and knowledge disciplines and justify the significance of their study.
5	Preparation of a review article
6	Identification of variables of a research study and their classification in terms of functions and level of measurement
7	Preparation of a sampling design given the objectives and research questions/hypotheses of a research study
8	Preparation of questionnaire for micro-level educational survey
9	Prepare 1 proposal on an identified research problem
10	Checking and removing plagiarism using Plagiarism Detection Software

Optimization Techniques (SIUSDS45)

Course Objective:

- Introduce optimization methodology as a valuable decision support tool.
- Help develop skills in building and solving optimization models for variety of engineering and Data Science related decision problems.
- Expose key mathematical concepts underlying various optimization models and algorithms.
- Provide hands-on experience with optimization software for solving and analyzing optimization models.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Formulate deterministic mathematical programs in various practical systems
CO2: Understand basic optimization techniques
CO3: Be able to interpret the results of a model and present the insights (sensitivity, duality)
CO4: Know the limitations of different solution methodology
CO5: Use software to solve problems

Unit	Contents	No. of Lectures
I	**Linear Programming Problem (LPP) :** Mathematical Model, Standard Form, Canonical Form, Cost Minimization & Profit Maximization Models, Graphical Polygon Corners Method, Graphical Iso-Profit/Iso-Cost Line Method, Analytical /Trial & Error Method, Simplex Method, Big-M Simplex Method, TwoPhase Simplex Method, Primal Vs Dual, Formulation of Dual from Primal, Comparison of Solutions of Primal & Dual	12
II	**Transportation Problem (TP) :** Mathematical Model, Cost Minimization & Profit Maximization Models of Balanced & Unbalanced Problems, North West Cost Method /Least Cost Method / Vogel's Approximation Method	12
III	**Assignment Problem (AP) :** Mathematical Model, Cost Minimization & Profit Maximization Models of Square Matrix & Non-Square Matrix Problems, Hungarian Method/ Reduced Matrix	12
IV	**Decision Making under Conditions of Certainty,Uncertainty & Risk :** Maximax, MaxiMin, MiniMax, Hurwicz, Laplace Criteria & Methods, EMV, EOL & EVPI Calculations, Incremental/ Marginal Analysis Method.	12

V	**Simulation Techniques :** Concepts & Applications, Monte Carlo Method, Problems of Simulation. **Decision Making under Conditions of Conflict - Game Theory :** Concepts & Applications, Simple Mathematical Models.	12

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Operations Research,	Prem Kumar Gupta & D S Hira	S Chand	---	---
2	Introduction to Mathematical Programming	Wayne L. Winston and Munirpallam Venkataramanan	---	4th Edition).	---

Internal Evaluation: 40 Marks

20 Marks	20 Marks
Class test	Assignment / Tutorial Work

Optimization Techniques Practical (SIUSDSP45)

List of Practical:

1	Simplex Method using R/ python/ scilab/ matlab.
2	North West Corner Method using R/ python/ scilab/ matlab.
3	Least Cost Method using R/ python/ scilab/ matlab.
4	Vogel's Approximation Method using R/ python/ scilab/ matlab.
5	Hungarian Method using R/ python/ scilab/ matlab.
6	**Decision Making under conditions of certainty / uncertainty:** i. Maximax ii. Maximin iii. Minimax
7	Monte Carlo Model Simulation using R/ python/ scilab/ matlab.
8	Decision Making under Conditions of Conflict - Game Theory **[Only Demonstration]**

Evaluation Scheme

I. Internal Exam-40 Marks

(i) Test– 20 Marks

20 marks Test – Duration 20 mins It will be conducted either using any open source learning management system such as Moodle (Modular object-oriented dynamic learning environment)Or a test based on an equivalent online course on the contents of the concerned course(subject)offered by or build using MOOC (Massive Open Online Course)platform.

(ii) 20 Marks - Active participation in routine class instructional deliveries Overall conduct as a responsible student, manners, skill in
Articulation, leadership qualities demonstrated through organizing co-curricular activities, etc.
Activities (Online Certification, International Certifications, Paper Presentation, Poster Presentation etc.)

II. External Examination- 60 Marks (i) Duration - 2 Hours.

(ii) Theory question paper pattern:-

All questions are compulsory.		
Question	Based on	Marks
Q.1	Unit I	12
Q.2	Unit II	12
Q.3	Unit III	12
Q.4	Unit IV	12
Q. 5	Unit V	12

- All questions shall be compulsory with internal choice within the questions.
- Each Question may be sub-divided into sub questions as a, b, c, d & e, etc & the allocation of Marks depends on the weightage of the topic.

III. Practical Examination – 250 marks (50 marks x 5 core papers)

- Each core subject carries 50 Marks: 40 marks + 05 marks (journal)+ 05 marks(viva)
- Minimum 75 % practical from each core subjects are required to be completed and written in the journal**.**

(Certified Journal is compulsory for appearing at the time of Practical Exam)
