SIES COLLEGE OF COMMERCE & ECONOMICS
AUTONOMOUS
DEPARTMENT OF DATA SCIENCE

Date of BOS meeting: 19th April, 2024

Name of BOS Chairperson:

Ms. Taranum Mohd. Salim Shaikh

Sr. No.	Heading	Particulars
1	Title of the course	M. Sc. (Data Science)
2	Eligibility for admission	A student who has completed B.Sc(Data Science), B.Sc(Information Technology),B.Sc(Computer Science), B.Sc(Mathematics), B.Sc(Statistics), B.Sc(Physics), B.Sc(Electronics) and B.C.A from any recognized university in India with minimum of 55 percent or equivalent C.G.P.A is eligible for this program.
3	Minimum percentage	55 %
4	Semesters	III & IV
5	Level	PG
6	Pattern	02 years & 04 semesters CBGS
7	To be implemented from	From Academic year 2024-25 in a progressive manner

SIES COLLEGE OF COMMERCE & ECONOMICS
(AUTONOMOUS)
(Affiliated to University of Mumbai)
RE-ACCREDITED GRADE "A" BY NAAC

BOARD OF STUDIES
DATA SCIENCE

(WITH EFFECT FROM THE ACADEMIC YEAR 2023-2024)

M. Sc (Data Science) – Part II

SEMESTER - III					
	Course Title				
Course Code	Theory	Credits	Course Code	Practical	Credits
MDC-MAJS3-501	Big Data Analytics	4	MDC-MJPS3-502	Big Data Analytics Practical	2
MDC-MAJS3-503	Data Science -II	4	MDC-MJPS3-504	Data Science -II Practical	2
Elective : Select Any one from the courses listed below along with corresponding practical course					
MDC-ELES3-505	Data Visualization	4	MDC-ELPS3-506	Data Visualization Practical	2
MDC-ELES3-507	Robotic Process Automation	4	MDC-ELPS3-508	Robotic Process Automation Practical	2
			MDC-RPS3-509	Research Project	4
	Total Theory Credits	**12**		Total Practical Credits	**10**
Total Credits for Semester III: 22					

SEMESTER – III

Big Data analytics

COURSE CODE: MDC-MAJS3-501 **COURSE CREDIT: 04**

Learning Objective:

The main goal of this course is to help students learn, understand, and practice big data analytics approaches, which include the conceptualization and summarization of big data and machine learning, and big data computing technologies.

Theory Component:

M. Sc (Data Science)	**Semester – III – MDC-MAJS3-501**
Course Name	**Big Data analytics**
Periods per week (1 Period is 60 minutes)	**4**
Credits (Theory + Internals)	**4**

Unit	Contents	No. of Lectures
I	**Introduction:** Introduction to Big Data, Big Data Characteristics, Types of Big Data, Traditional Versus Big Data Approach, Technologies Available for Big Data, Infrastructure for Big Data, Use of Data Analytics, Big Data Challenges, Desired Properties of a Big Data System, Case Study of Big Data Solutions.	12
II	**Analytical Theory and Methods:** Clustering and Associated Algorithms, Association Rules, Apriori Algorithm, Candidate Rules, Applications of Association Rules, Validation and Testing, Diagnostics Classification, Decision Trees, Time Series Analysis, Text Analysis, Steps, Text Analysis Example, Collecting Raw Text, Representing Text, Term Frequency-Inverse Document Frequency (TFIDF), Categorizing Documents by Topics, Determining Sentiments.	12
III	**Hadoop:** Introduction, what is Hadoop? Core Hadoop Components, Operating System for Big Data, Concepts, Hadoop Architecture, Hadoop Ecosystem, Hadoop Limitations, Recommendation Systems.	12

IV	**NoSQL:** What is NoSQL? NoSQL Business Drivers, NoSQL Case Studies, NoSQL Data Architectural Patterns, Variations of NoSQL Architectural Patterns, Using NoSQL to Manage Big Data **Map Reduce**: MapReduce and The New Software Stack, MapReduce, Algorithms Using MapReduce. **YARN: Introduction** to yarn and components of yarn	**12**
V	**Introduction to HBase (Column family stores):** HBase history, what is HBase? Difference between HBase and HDFS, HBase - features, applications, Storage mechanism, Architecture, reading and writing of data., compaction process, file types. Introduction to Pig: What is Pig? Advantages, Components of Pig. Hive- What is Hive? Hive and RDBMS, Features, architecture, components of Hive.	**12**

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Big Data Analytics	Radha Shankarmani	Wiley	Second	2016
2	Big Data and Analytics	Subhashini Chellappan Seema Acharya	Wiley	First	2015
3	Big Data Analytics with R and Hadoop	Vignesh Prajapati	Packt	First	2013
4	Practical Big data Analytics	Nataraj Dasgupta	Packt	First	2018
5	Big Data Analytics	Anuradha Bhatia			

Data Science – II

COURSE CODE: MDC-MAJS3-503 **COURSE CREDIT: 04**

Learning Objective: Learners can apply various modeling and data analysis techniques to the solution of real-world business problems, impart findings, and effectively present results using data visualization techniques

Theory Component:

M. Sc (Data Science)	**Semester – III – MDC-MAJS3-503**
Course Name	**Data Science – II**
Periods per week (1 Period is 60 minutes)	**4**
Credits (Theory + Internals)	**4**

Unit	Contents	No. of Lectures
I	**Exploratory Data Analysis Checklist:** Formulate your question, read in your data, Check the packaging, run str (), Look at the top and the bottom of your data, check your —n‖s, validate with at least one external data source, Try the easy solution first, challenge your solution, Follow up questions **Principles of Analytic Graphics:** Show comparisons, show causality, mechanism, explanation, systematic structure, show multivariate data, integrate evidence, Describe and document the evidence, Content, Content, Content	12
II	**Exploratory Graphs:** Characteristics of exploratory graphs, Air Pollution in the United States, Getting the Data, Simple Summaries: One Dimension, Five Number Summary, Boxplot, Histogram, Overlaying Features, Barplot, Simple Summaries: Two Dimensions and Beyond, Multiple Boxplots, Multiple Histograms, Scatterplots, Scatterplot - Using Color, Multiple Scatterplots **Plotting Systems:** The Base Plotting System, The Lattice System, The ggplot2 System **Graphics Devices:** The Process of Making a Plot, How Does a Plot Get Created? Graphics File Devices, Multiple Open Graphics Devices, Copying Plots	12
III	**The Base Plotting System:** Base Graphics, Simple Base Graphics, Some Important Base Graphics Parameters, Base Plotting Functions, Base Plot with Regression Line, Multiple Base Plots **Plotting and Color in R:** Colors 1, 2, and 3, Connecting colors with data, Color Utilities in R, colorRamp(), colorRampPalette(), RColorBrewer Package, Using the RColorBrewer palettes, the smoothScatter () function, Adding transparency	12

IV	**Hierarchical Clustering:** Hierarchical clustering, how do we define close? Example: Euclidean distance, Example: Manhattan distance, Example: Hierarchical clustering, Prettier dendrograms, Merging points: Complete, Merging points: Average, Using the heatmap() function, Notes and further resources, K-Means Clustering, Illustrating the K-means algorithm, Stopping the algorithm, Using the kmeans () function, Building heatmaps from K-means solutions, Notes and further resources **Dimension Reduction:** Matrix data, Patterns in rows and columns, Related problem, SVD and PCA, Unpacking the SVD: *u* and *v*, SVD for data compression, Components of the SVD - Variance explained, Relationship to principal components, what if we add a second pattern? Dealing with missing values, Example: Face data, Notes and further resource	**12**
V	**The ggplot2 Plotting System - Part 1:** The Basics: qplot(), Before You Start: Label Your Data, ggplot2 ―Hello, world!‖, Modifying aesthetics, Adding a geom, Histograms, Facets, Case Study: MAACS Cohort **The ggplot2 Plotting System - Part 2:** Basic Components of a ggplot2 Plot, Example: BMI, PM2.5, Asthma, Building Up in Layers, First Plot with Point Layer, Adding More Layers: Smooth, Adding More Layers: Facets, Modifying Geom Properties, Modifying Labels, Customizing the Smooth, Changing the Theme, More Complex Example, A Quick Aside about Axis Limits, Resources **Data Analysis Case Study: Changes in Fine Particle Air Pollution in the U.S.:** Synopsis, Loading and Processing the Raw Data, Results	**12**

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Exploratory Data Analysis	Roger D. Peng		1st	2016
2	R Programming for Data Science	Roger D Peng		1st	2015
3	Data Science from Scratch	Joel Grus	O'Reilly Media, Inc.	2nd	2019
4	R for Data Science	Hadley Wickham, Garrett Grolemund	O'Reilly Media, Inc.	1st	2016

ELECTIVES

Data Visualization

COURSE CODE: MDC-ELES3-505 **COURSE CREDIT: 04**

Learning Objective:

- To apply the functionality of the various data visualization tools and techniques
- To understand visual perception, visual representation of data
- To understand and apply various classification and prediction techniques using tools.
- To study and apply Visualization of groups, trees, graphs, clusters, networks on data set.
- To understand Mining of Object, Spatial, Multimedia, Text and Web Data.

Theory Component:

M. Sc (Data Science)	**Semester – III – MDC-ELES3-505**
Course Name	**Data Visualization**
Periods per week (1 Period is 60 minutes)	**4**
Credits (Theory + Internals)	**4**

Unit	Contents	No. of Lectures
I	Introduction of visual perception, visual representation of data, Gestalt principles, information overloads, Design principles Categorical, time series, and statistical data graphics.	**12**
II	Creating visual representations, visualization reference model, visual mapping, visual analytics, Design of visualization applications.	**12**
III	Classification of visualization systems, Interaction and visualization techniques misleading, Visualization of one, two and multi-dimensional data, text and text documents.	**12**
IV	Visualization of groups, trees, graphs, clusters, networks, software, Metaphorical visualization.	**12**
V	Visualization of volumetric data, vector fields, processes and simulations, Visualization of maps, geographic information, GIS systems, collaborative visualizations, evaluating visualizations.	**12**

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Interactive Data Visualization: Foundations, Techniques, and Applications.	Ward, Grinstein Keim	A K Peters/CRC Press	Second	2015
2	The Visual Display of Quantitative Information	E. Tufte	Graphics Press	Second	2001

Robotic Process Automation

COURSE CODE: MDC-ELES3-507 **COURSE CREDIT: 04**

Learning Objective:

- To make the students aware about the automation today in the industry.
- To make the students aware about the tools used for automation.
- To help the students automate a complete process.

Theory Component:

M. Sc (Data Science)	**Semester – III – MDC-ELES3-507**
Course Name	**Robotic Process Automation**
Periods per week (1 Period is 60 minutes)	**4**
Credits (Theory + Internals)	**4**

Sr. No	Modules/Units	No of Lectures
1.	**Robotic Process Automation:** Scope and techniques of automation, About UiPath **Record and Play:** UiPath stack, Downloading and installing UiPath Studio, Learning UiPath Studio, Task recorder, Step-by-step examples using the recorder.	12
2.	**Sequence, Flowchart, and Control Flow:** Sequencing the workflow, Activities, Control flow, various types of loops, and decision making, Step-by-step example using Sequence and Flowchart, Step-by-step example using Sequence and Control flow **Data Manipulation:** Variables and scope, Collections, Arguments – Purpose and use, Data table usage with examples, Clipboard management, File operation with step-by-step example, CSV/Excel to data table and vice versa (with a step-by-step example)	12
3.	**Taking Control of the Controls :** Finding and attaching windows, Finding the control, Techniques for waiting for a control, Act on controls – mouse and keyboard activities, Working with UiExplorer, Handling events, Revisit recorder, Screen Scraping, When to use OCR, Types of OCR available, How to use OCR, Avoiding typical failure points **Tame that Application with Plugins and Extensions:** Terminal plugin, SAP automation, Java plugin, Citrix automation, Mail plugin, PDF plugin, Web integration, Excel and Word plugins, Credential management, Extensions – Java, Chrome, Firefox, and Silverlight	12

4.	**Handling User Events and Assistant Bots:** What are assistant bots?, Monitoring system event triggers, Hotkey trigger, Mouse trigger, System trigger, Monitoring image and element triggers, An example of monitoring email, Example of monitoring a copying event and blocking it, Launching an assistant bot on a keyboard event **Exception Handling, Debugging, and Logging:** Exception handling, Common exceptions and ways to handle them, Logging and taking screenshots, Debugging techniques, Collecting crash dumps, Error reporting	12
5.	**Managing and Maintaining the Code:** Project organization, Nesting workflows, Reusability of workflows, Commenting techniques, State Machine, When to use Flowcharts, State Machines, or Sequences, Using config files and examples of a config file, Integrating a TFS server **Deploying and Maintaining the Bot:** Publishing using publish utility, Overview of Orchestration Server, Using Orchestration Server to control bots, Using Orchestration Server to deploy bots, License management, Publishing and managing updates	12

REFERENCE BOOKS:

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Learning Robotic Process Automation	Alok Mani Tripathi	Packt	1st	2018
2.	Robotic Process Automation Tools, Process Automation and their benefits: Understanding RPA and Intelligent Automation	Srikanth Merianda	Createspace Independent Publishing	1st	2018
3.	The Simple Implementation Guide to Robotic Process Automation (Rpa): How to Best Implement Rpa in an Organization	Kelly Wibbenmeyer	iUniverse	1st	2018

Practical Component: (SEMESTER III)

Big Data analytics Practical

COURSE CODE: MDC-MJPS3-502 **COURSE CREDIT: 02**

M. Sc (Data Science)	**Semester – III – MDC-MJPS3-502**
Course Name	**Big Data analytics Practical**
Periods per week (1 Period is 60 minutes)	**4**
Credits	**2**

List of Practical:

1	Installation of HADOOP.
2	Implement the following file management tasks in Hadoop System (HDFS): Adding files and directories, retrieving files, Deleting files.
3	Basic CRUD operations in MongoDB
4	To understand the overall programming architecture using Map Reduce API and implement programs related to MapReduce
5	Implement clustering and associated algorithms
6	Implement Linear Regression
7	Implement Bloom Filters for filter on Stream Data
8	Implement Time Series
9	Creating the HDFS tables and loading them in Hive and learn joining of tables in Hive
10	To perform NoSQL database using mongodb to create, update and insert.
11	Implement a simple recommender system
12	Assignment problem.
13	PERT/CPM problem.

Practical Component: (SEMESTER III)

Data Science – II Practical

COURSE CODE: MDC-MJPS3-504 **COURSE CREDIT: 02**

M. Sc (Data Science)	**Semester – III - MDC-MJPS3-504**
Course Name	**Data Science – II Practical**
Periods per week (1 Period is 60 minutes)	**4**
Credits	**2**

List of Practicals:

1	Find out the age of Abalone from physical measurements. Use Regression Models. Use the data set **abalone.data.csv**
2	Predict student's knowledge level. Use Classification/Clustering Models. Use the data set **Data_User_Modeling_Dataset_Hamdi Tolga KAHRAMAN.xls**
3	Can you estimate location from WIFI Signal Strength Use Classification Models Use the data set **wifi_localization.txt**
4	Predict acceptability of a car. Use Classification Models. Use the data set **car.data**
5	Predict total number of demand of orders. Use Regression Models. Use the data set **Daily_Demand_Forecasting_Orders.csv**
6	Forecast pollution level of a city. Use Regression Models. Use the data set **PRSA_data_2010.1.1-2014.12.31.csv**
7	Will the patient survive for at least one year after a heart attack Use Classification Models Use the data set **echocardiogram.data**
8	Predict which stock will provide greatest rate of return. Use Classification/Clustering/ Regression Models. Use the data set **dow_jones_index.data**

Practical Component: (SEMESTER III)

Data Visualization Practical (Electives)

COURSE CODE: MDC-ELPS3-506 **COURSE CREDIT: 02**

M. Sc (Data Science)	**Semester – III - MDC-ELPS3-506**
Course Name	**Data Visualization - Practical (Electives)**
Periods per week (1 Period is 60 minutes)	**4**
Credits	**2**

List of Practicals:

1	Demonstrate a nonobvious insight gleaned from the data, or to make a particular point. You can stick with a single chart or other type of visualization, or you can use multiple displays that together tell a story. To create your visualization(s) you can use the simple tools - spreadsheet graphing tools (e.g., Google Sheets, Excel), Tableau tool, or any other method of creating a visual point or story from the data.
2	Demonstrate Time series and statistical data graphics using visualization tool.
3	Generating visualizations of map-based data.
4	Finding data There are many sources of freely downloadable data. Locate a relevant data set online. Here are some places to start: • LION: http://www.nyc.gov/html/dcp/html/bytes/applbyte.shtml •Newman Library: http://guides.newman.baruch.cuny.edu/nyc_data • NYC Open Data: http://data.cityofnewyork.us • Wiki: https://wiki.gephi.org/index.php/Datasets Demonstrate temporal component, showing change over time.
5	Demonstrate visualization of one, two and multi-dimensional data.
6	Visualizing tenure, monthly charges, total charges, and other individual columns using a scatter plot
7	Demonstrate visualization of text and text documents.
8	Create a Map view with appropriate data set using tableau.
9	Demonstrate Metaphorical visualization.
10	Demonstrate visualization of groups, trees, graphs, clusters.
11	Demonstrate collaborative visualization

Practical Component:
(SEMESTER III)

Robotic Process Automation
Practical (Electives)

COURSE CODE: MDC-ELPS3-508 **COURSE CREDIT: 02**

M. Sc (Data Science)	**Semester – III - MDC-ELPS3-508**
Course Name	**Robotic Process Automation - Practical (Electives)**
Periods per week (1 Period is 60 minutes)	**4**
Credits	**2**

List of Practicals:

1.	a. Create a simple sequence based project.
	b. Create a flowchart-based project.
	c. Create an UiPath Robot which can empty a folder in Gmail solely on basis of recording.
2.	a. Automate UiPath Number Calculation (Subtraction, Multiplication, Division of numbers).
	b. Create an automation UiPath project using different types of variables (number, datetime, Boolean, generic, array, data table)
3.	a. Create an automation UiPath Project using decision statements.
	b. Create an automation UiPath Project using looping statements.
4.	a. Automate any process using basic recording.
	b. Automate any process using desktop recording.
	c. Automate any process using web recording.
5.	a. Consider an array of names. We have to find out how many of them start with the letter "a". Create an automation where the number of names starting with "a" is counted and the result is displayed.
6.	a. Create an application automating the read, write and append operation on excel file.
	b. Automate the process to extract data from an excel file into a data table and vice versa
7.	a. Implement the attach window activity.
	b. Find different controls using UiPath.
	c. Demonstrate the following activities in UiPath: i. Mouse (click, double click and hover) ii. Type into iii. Type Secure text
8.	a. Demonstrate the following events in UiPath: i. Element triggering event ii. Image triggering event iii. System Triggering Event
	b. Automate the following screen scraping methods using UiPath i. Full Test ii. Native iii. OCR

	c. Install and automate any process using UiPath with the following plug-ins: i. Java Plugin ii. Mail Plugin iii. PDF Plugin iv. Web Integration

Research Project

COURSE CODE: MDC-RPS3-509 **COURSE CREDIT: 04**

The Research Project documentation and implementation and Viva Voce details are given in **Appendix 1**.

DEPARTMENT OF DATA SCIENCE
PROPOSED SCHEME OF EXAMINATION

Evaluation Scheme

Internal Evaluation (40 Marks)

The internal assessment marks shall be awarded as follows:

1. 30 marks (Any one of the following):
    - **a.** Written Test or
    - **b.** SWAYAM (Advanced Course) of minimum 20 hours and certification exam completed or
    - **c.** NPTEL (Advanced Course) of minimum 20 hours and certification exam completed or
    - **d.** Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy and the like)
    - **e.** One certification marks shall be awarded one course only. For four courses, the students will have to complete four certifications.
    - **f.** Research paper publication
2. 10 marks
    - **a.** Assignments/ Group discussions/ Debates/ Quiz/ Subject specific case study/ Mini Project/ Presentation/ Field work/ Program implementation/ any other

External Examination: (60 marks)

	All questions are compulsory	
Q1	(Based on Unit 1) Attempt *any two* of the following:	12
a.		
b.		
c.		
d.		
Q2	(Based on Unit 2) Attempt *any two* of the following:	12
Q3	(Based on Unit 3) Attempt *any two* of the following:	12
Q4	(Based on Unit 4) Attempt *any two* of the following:	12
Q5	(Based on Unit 5) Attempt *any two* of the following:	12

Practical Evaluation (50 marks)

A Certified copy journal is essential to appear for the practical examination.

1.	Practical Question 1	20
2.	Practical Question 2	20
3.	Journal	5
4.	Viva Voce	5

OR

1.	Practical Question	40
2.	Journal	5
3.	Viva Voce	5

Appendix – 1

Research Project Documentation and Implementation Viva-voce (Semester III)

Goals of the course Project Documentation and Viva-Voce

The student should:

- be able to apply relevant knowledge and abilities, within the main field of study, to a given problem
- within given constraints, even with limited information, independently analyse and discuss complex inquiries/problems and handle larger problems on the advanced level within the main field of study
- reflect on, evaluate and critically review one's own and others' scientific results
- be able to document and present one's own work with strict requirements on structure, format, and language usage
- be able to identify one's need for further knowledge and continuously develop one's own knowledge

To start the project:

- Start thinking early in the programme about suitable projects.
- Read the instructions for the project.
- Attend and listen to other student´s final oral presentations.
- Look at the finished reports.
- Talk to senior master students.
- Attend possible information events (workshops / seminars / conferences etc.) about therelated topics.

Application and approval:

- Read all the detailed information about project.
- Finalise finding a place and supervisor.
- Check with the coordinator about subject/project, place and supervisor.
- Write the project proposal and plan along with the supervisor.
- Fill out the application together with the supervisor.
- Hand over the complete application, proposal and plan to the coordinator.
- Get an acknowledgement and approval from the coordinator to start the project.

During the project:

- Search, gather and read information and literature about the theory.
- Document well the practical work and your results.
- Take part in seminars and the running follow-ups/supervision.
- Think early on about disposition and writing of the final report.
- Discuss your thoughts with the supervisor and others.
- Read the SOP and the rest you need again.

- Plan for and do the mid-term reporting to the coordinator/examiner.
- Do a mid-term report also at the work-place (can be a requirement in some work-places).

- Write the first draft of the final report and rewrite it based on feedback from the supervisor and possibly others.
- Plan for the final presentation of the report.

Finishing the project:

- Finish the report and obtain an OK from the supervisor.
- Ask the supervisor to send the certificate and feedback form to the coordinator.
- Attend the pre-final oral presentation arranged by the Coordinator.
- Rewrite the final report again based on feedback from the opponents and possibly others.
- Prepare a title page and a popular science summary for your report.
- Send the completed final report to the coordinator (via plagiarism software)
- Rewrite the report based on possible feedback from the coordinator.
- Appear for the final exam.

Project Proposal/research plan

- The student should spend the first 1-2 weeks writing a 1-2 pages project plan containing:
- Short background of the project
- Aims of the project
- Short description of methods that will be used
- Estimated time schedule for the project
- The research plan should be handed in to the supervisor and the coordinator.
- Writing the project plan will help you plan your project work and get you started in finding information and understanding of methods needed to perform the project.

Project Documentation

- The documentation should contain:
- Introduction - that should contain a technical and social (when possible) motivation of the project topic.
- Description of the problems/topics.
- Status of the research/knowledge in the field and literature review.
- Description of the methodology/approach. (The actual structure of the chapters here depends on the topic of the documentation.)
- Results - must always contain analyses of results and associated uncertainties.
- Conclusions and proposals for the future work.
- Appendices (when needed).
- Bibliography - references and links.

SIES COLLEGE OF COMMERCE & ECONOMICS
AUTONOMOUS
DEPARTMENT OF DATA SCIENCE

Date of BOS meeting: 26th October, 2024

Name of BOS Chairperson: Ms. Taranum Mohd. Salim Shaikh

Sr. No.	Heading	Particulars
1	Title of the course	M. Sc. (Data Science)
2	Eligibility for admission	1. A student who has completed B.Sc(Information Technology),B.Sc(Computer Science), B.Sc(Mathematics), B.Sc(Statistics), B.Sc(Physics), B.Sc(Electronics) and B.C.A from any recognized university in India with minimum of 55 percent or equivalent C.G.P.A is eligible for this program. 2. It is mandatory that all students should have done Calculus and Linear Algebra course during their graduation. 3. An entrance test(online) will be conducted on General and Logical Aptitude (60 marks) The selected students will go through an interview process (40 marks)
3	Minimum percentage	55 %
4	Semesters	III & IV
5	Level	PG
6	Pattern	02 years & 04 semesters CBGS
7	To be implemented from	From Academic year 2024-25 in a progressive manner

SIES COLLEGE OF COMMERCE & ECONOMICS
(AUTONOMOUS)
(Affiliated to University of Mumbai)
RE-ACCREDITED GRADE "A" BY NAAC

BOARD OF STUDIES
DATA SCIENCE

(WITH EFFECT FROM THE ACADEMIC YEAR 2024-2025)

M. Sc (Data Science) – Part II

SEMESTER - IV					
	Course Title				
Course Code	Theory	Credits	Course Code	Practical	Credits
MDC-MAJS4-501	Deep Learning	4	MDC-MJPS4-502	Deep Learning Practical	2
MDC-MAJS4-503	Data Storage and Management	4	MDC-MJPS4-504	Data Storage and Management Practical	2
Elective : Select Any one from the courses listed below along with corresponding practical course					
MDC-ELES4-505	Web and Social Network Data Analytics	3	MDC-ELPS4-506	Web and Social Network Data Analytics Practical	1
MDC-ELES4-507	Natural Language Processing	3	MDC-ELPS4-508	Natural Language Processing Practical	1
			MDC-RPS4-509	Research Project	6
	Total Theory Credits	**11**		Total Practical Credits	**11**
Total Credits for Semester IV: 22					

SEMESTER – IV

Deep Learning

COURSE CODE: MDC-MAJS4-501 **COURSE CREDIT: 04**

Learning Objective:

- To know importance of deep learning
- To acquire knowledge on the basics of neural networks.
- To implement neural networks using computational tools for variety of problems.
- To explore various deep learning algorithms

Theory Component:

M. Sc (Data Science)	**Semester – IV – MDC-MAJS4-501**
Course Name	**Deep Learning**
Periods per week (1 Period is 60 minutes)	**4**
Credits (Theory + Internals)	**4**

Unit	Contents	No. of Lectures
I	**Introduction**: Introduction, History, Capabilities, the perceptron **Neural network learning**: Back-propagation, practical network training, Auto encoders, Batch-normalization: Why does it work? Overfitting and generalization, Perceptron and MLP, FFN, Back propagation, **Activation Functions**: Sigmoid, ReLU, Hyperbolics FNS, Softmax, **Artificial Neural Networks**: Introduction, Perceptron Training Rule, Gradient Decent Rule, **Gradient Decent and Back propagation**: Gradient Decent, Stochastic Gradient Decent, Back propagation, Some problems in ANN, **Optimization and Regularization**: Overfitting and Capacity, Cross Validation, Feature Selection, Regularization, Hyperparameters	**12**
II	**Convolution Neural Network**: Introduction to CNNs, Convolution, Correlation, Filtering, CNN Architectures, Detection and Segmentation, Image classification, Text classification, Image classification and hyper-parameter tuning, Advanced CNNs for computer vision	**12**
III	**Recurrent Neural Networks**: Introduction to RNNs, Unfolded RNNs, Seq2Seq RNNs, LSTM, RNN applications, Building recurrent NN, Long Short – Term Memory, Time Series Forecasting	**12**
IV	**Deep Unsupervised Learning**: Stacked auto-encoders and semi-supervised learning, **Regularization:** Dropout and Batch normalization	**12**
V	**Advanced Deep architectures**: Advanced RNN: LSTM, GRU, Generative Adversarial Networks (GANs), Advanced GANs, Emerging NN architectures	**12**

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Deep Learning	Ian Goodfellow, Yoshua Bengio, Aaron Courville	MIT Press		2020
2	Neural Networks and Deep Learning	Michael Nielsen	Determination Press		2015
3	Deep Learning with Python	Francois Chollet	Manning Publications	First	2017
4	Pattern Recognition and Machine Learning	Bishop, C.M	Springer		2021
5	Artificial Neural Networks	Yegnanarayana, B.	PHI Learning		2009
6	Matrix Computations	Golub, G.H and Van Loan C.F.	JHU Press		2013
7	Neural Networks: A Classroom Approach	Satish Kumar	Tata McGraw-Hill Education		2022

Data Storage and Management

COURSE CODE: MDC-MAJS4-503 **COURSE CREDIT: 04**

Learning Objective:

- Understand the types of storage systems.
- Utilize redundant array of independent disks (RAID) technologies effectively
- Understanding the positioning of data at various level of memory hierarchy.
- Learning Distributed data base system, Mango DB, Storage architecture, AN.

Theory Component:

M. Sc (Data Science)	**Semester – IV – MDC-MAJS4-503**
Course Name	**Data Storage and Management**
Periods per week (1 Period is 60 minutes)	**4**
Credits (Theory + Internals)	**4**

Unit	Contents	No. of Lectures
I	Storage Media and Technologies – Magnetic, Optical and Semiconductor Media, Techniques for read/write Operations, Issues and Limitations.	**12**
II	Usage and Access – Positioning in the Memory Hierarchy, Hardware and Software Design for Access, Performance issues. Distributed Database Patterns— Distributed Relational Databases- Non-relational Distributed Databases- MongoDB - Sharing and Replication- HBase-	**12**
III	Cassandra Consistency Models— Types of Consistency- Consistency MongoDB- HBase Consistency- Cassandra Consistency. Large Storages – Hard Disks, Networked Attached Storage, Scalability issues, Networking issues.	**12**
IV	Storage Architecture - Storage Partitioning, Storage System Design, Caching, Legacy Systems.	**12**
V	Storage Area Networks – Hardware and Software Components, Storage Clusters/Grids. Storage QoS–Performance, Reliability, and Security issues, storage appliances. Network and web security: Network Security: Network Concepts, Threats in Networks, Network Security Controls. Web Security Requirements, Secure Socket Layer (SSL), Transport Layer Security (TLS), Secure Electronic Transaction (SET).	**12**

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	The Complete Guide to Data Storage Technologies for Network-centric Computing	Franklyn E. Dailey Jr.	Computer Technology Research Corporation	First	1998
2	Data Storage Networking	Nigel Poulton	Sybex	Third	2018

ELECTIVES

Web and Social Network Data Analytics

COURSE CODE: MDC-ELES4-505 **COURSE CREDIT: 03**

Learning Objective:

- To introduce the concepts of Web and Information Retrieval and Web Mining in Social Network.
- To study the basic concepts of Social Network Analysis.
- To interpret social networks through mathematical representation.
- To analyze relations, descriptive measures and models to overview research questions related to Social Networks.

Theory Component:

M. Sc (Data Science)	**Semester – IV – MDC-ELES4-505**
Course Name	**Web and Social Network Data Analytics**
Periods per week (1 Period is 60 minutes)	**3**
Credits (Theory + Internals)	**3**

Unit	Contents	No. of Lectures
I	**Information Retrieval and Web Search:** Basic Concepts, Information Retrieval Models, Text and Web Page Pre-Processing, Inverted Index and Its Compression, Latent Semantic Indexing, Web Search, Meta-Search, Web Spamming.	**12**
II	**Social Network Analysis**: Co-Citation and Bibliographic Coupling, PageRank, HITS Algorithm, Community Discovery **Web Crawling**: A Basic Crawler Algorithm, Implementation Issues, Universal Crawlers, Topical Crawlers, Crawler Ethics and Conflicts.	**12**
III	**Structured Data Extraction: Wrapper Generation:** Preliminaries, Wrapper Induction, Instance-Based Wrapper Learning, Automatic Wrapper Generation, String Matching and Tree Matching, Multiple Alignment, Flat Data Records, Nested Data Records, Extraction Based on Multiple Pages.	**12**
IV	**Opinion Mining and Sentiment Analysis:** The Problem of Opinion Mining, Document Sentiment Classification, Sentence Subjectivity and Sentiment Classification, Aspect-Based Opinion Mining, Mining Comparative Opinions, Opinion Search and Retrieval, Opinion Spam Detection.	**12**
V	**Web Usage Mining:** Data Collection and Pre-Processing, Data Modeling, Discovery and Analysis of Web Usage Patterns, Recommender Systems and Collaborative Filtering, Query Log Mining.	**12**

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data	Bing Liu	Springer	Third	2021
2	Mining the Social Web	Matthew A. Russell and Mikhail Klassen	O'Reilly	Fourth	2022
3	Analyzing Social Networks.	Stephen P. Borgatti	SAGE	Second	2017

Natural Language Processing
COURSE CODE: MDC-ELES4-507 **COURSE CREDIT: 03**

Learning Objective:

- The prime objective of this course is to introduce the students to the field of Language Computing and its applications ranging from classical era to modern context.
- To provide understanding of various NLP tasks and NLP abstractions such as Morphological analysis, POS tagging, concept of syntactic parsing, semantic analysis etc.
- To provide knowledge of different approaches/algorithms for carrying out NLP tasks.
- To highlight the concepts of Language grammar and grammar representation in Computational Linguistics.

Theory Component:

M. Sc (Data Science)	**Semester – IV – MDC-ELES4-507**
Course Name	**Natural Language Processing**
Periods per week (1 Period is 60 minutes)	**3**
Credits (Theory + Internals)	**3**

Sr. No	**Modules/Units**	**No of Lectures**
1.	Introduction to NLP, Brief history, Working of NLP NLP applications: Speech to Text(STT), Text to Speech(TTS), Text Summarization, Text classification, Sentiment Analysis, Grammar/Spell Checkers, NL tasks: Segmentation, Chunking, tagging, NER, Parsing, Word Sense Disambiguation, NL Generation, Sentiment Analysis, Text Entailment, Cross Lingual Information Retrieval (CLIR)	12
2.	Text Processing Challenges, Segmentation: word level(Tokenization), Sentence level. Regular Expression and Automata Morphology, Types, Survey of English and Indian Languages Morphology, Morphological parsing FSA and FST, Porter stemmer, Rule based and Paradigm based Morphology, Human Morphological Processing, Machine Learning approaches	12

3.	Word Classes ad Part-of-Speech tagging(POS), survey of POS tagsets, Rule based approaches (ENGTOWL), Stochastic approaches(Probabilistic, N-gram and HMM), Evaluation metrics: Precision/Recall/F-measure, error analysis.	12
4.	NL parsing basics, approaches: TopDown, BottomUp, Overview of Grammar Formalisms: constituency and dependency school, Grammar notations CFG, LFG, PCFG, LTAG, Feature-Unification, overview of English CFG, Indian Language Parsing in Paninian Karaka Theory, Probabilistic parsing, Dependency	12

	Parsing: Covington algorithm, MALT parser, MST parser.	
5.	Concepts and issues in NL, Theories and approaches for Semantic Analysis, Meaning Representation, word similarity, Lexical Semantics, word senses and relationships, WordNet (English and IndoWordnet), Word Sense Disambiguation: Lesk Algorithm Walker's algorithm, Coreferences Resolution: Anaphora, Cataphora	12

REFERENCE BOOKS:

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Handbook of Natural Language Processing	Indurkhya, N., & Damerau, F. J.	CRC Press Taylor and Francis Group	2nd	2010
2.	Speech and Language Processing	Martin, J. H., & Jurafsky, D.	Pearson Education India	2nd	2013
3.	Foundations of Statistical Natural Language Processing	Manning, Christopher and Heinrich, Schutze	MIT Press	1st	1997
4.	Natural Language Processing With Python	Steven Bird, Edward Loper	O'Reilly Media	2nd	2016
5.	Video Links				

Practical Component: (SEMESTER IV)

Deep Learning - Practical

COURSE CODE: MDC-MJPS4-502 **COURSE CREDIT: 02**

M. Sc (Data Science)	**Semester – IV - MDC-MJPS4-502**
Course Name	**Deep Learning - Practical**
Periods per week (1 Period is 60 minutes)	**4**
Credits	**2**

List of Practicals:

1	Demonstrate Gradient descent and the back-propagation algorithm.
2	Implement RelU Heuristics for avoiding bad local minima.
3	Demonstrate Image segmentation, object detection.
4	Demonstrate automatic image captioning, Image generation with Generative adversarial networks.
5	Demonstrate video to text with LSTM model.
6	Demonstrate Neural Summarization using NLP
7	Demonstrate similar question detection using NLP.
8	Demonstrate Auto encoders using Deep Unsupervised Learning.
9	Demonstrate Variational Auto encoders.
10	Demonstrate Dialogue topic tracking.

Practical Component: (SEMESTER IV)

Data Storage and Management - Practical

COURSE CODE: MDC-MJPS4-504 **COURSE CREDIT: 02**

M. Sc (Data Science)	**Semester – IV - MDC-MJPS4-504**
Course Name	**Data Storage and Management - Practical**
Periods per week (1 Period is 60 minutes)	**4**
Credits	**2**

List of Practicals:

1	Demonstrate the Read/Write time of data from various storage devices (Pen Drive, HDD, CD/DVD)
2	Demonstrate the Usage and Access of data positioned at various level of memory hierarchy.
3	Build an application on private cloud.
4	Demonstrate any Cloud data storage Monitoring tool.
5	Implement FOSS-Cloud Functionality - VDI (Virtual Desktop Infrastructure)
6	Demonstrate Distributed Relational Databases
7	Demonstrate Non-relational Distributed Databases
8	Demonstrate parallel data base in peer-to-peer environment.
9	Implement FOSS-Cloud Functionality - VSI Software as a Service (SaaS).
10	Explore Working of the following with Virtual Machines- a. VM Lifecycle b. Creating VMs c. Accessing VMs d. Assigning VMs to Hosts
11	Explore Service Offerings, Disk Offerings, Network Offerings and Templates – In open-source Cloud technology

Practical Component: (SEMESTER IV)

Web and Social Network Data Analytics - Practical

COURSE CODE: MDC-ELPS4-506 **COURSE CREDIT: 01**

M. Sc (Data Science)	**Semester – IV - MDC-ELPS4-506**
Course Name	**Web and Social Network Data Analytics - Practical**
Periods per week (1 Period is 60 minutes)	**2**
Credits	**1**

List of Practicals: (To be implemented using any of the web mining tools)

1	Page Rank Algorithm
2	Weighted Page Rank Algorithm
3	HITS Algorithm
4	Crawler Algorithms
5	Structured Data Extraction through Wrapper Generation
6	Opinion Search and Retrieval
7	Sentiment Analysis
8	Web Content Mining
9	Web Structure Mining
10	Case Studies: Google, Facebook, Twitter, Instagram

Practical Component: (SEMESTER IV)

Natural Language Processing - Practical

COURSE CODE: MDC-ELPS4-508 **COURSE CREDIT: 01**

M. Sc (Data Science)	Semester – IV - MDC-ELPS4-508
Course Name	Natural Language Processing- Practical
Periods per week (1 Period is 60 minutes)	2
Credits	1

List of Practicals:

10 practicals covering the entire syllabus must be performed. The detailed list of practical will be circulated later in the official workshop.

Project Implementation and Viva

COURSE CODE: MIT-RPS4-509 **COURSE CREDIT: 06**

The project dissertation and Viva Voce details are given in **Appendix 1**.

DEPARTMENT OF DATA SCIENCE

PROPOSED SCHEME OF EXAMINATION

Evaluation Scheme

Internal Evaluation (40 Marks)

The internal assessment marks shall be awarded as follows:

1. **30 marks (Any one of the following):**
   - **a.** Written Test or
   - **b.** SWAYAM (Advanced Course) of minimum 20 hours and certification exam completed or
   - **c.** NPTEL (Advanced Course) of minimum 20 hours and certification exam completed or
   - **d.** Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy and the like)
   - **e.** One certification marks shall be awarded one course only. For four courses, the students will have to complete four certifications.
   - **f.** Research paper publication
2. **10 marks**
   - **a.** Assignments/ Group discussions/ Debates/ Quiz/ Subject specific case study/ Mini Project/ Presentation/ Field work/ Program implementation/ any other

External Examination: (60 marks)

	All questions are compulsory	
Q1	(Based on Unit 1) Attempt *any two* of the following:	12
a.		
b.		
c.		
d.		
Q2	(Based on Unit 2) Attempt *any two* of the following:	12
Q3	(Based on Unit 3) Attempt *any two* of the following:	12
Q4	(Based on Unit 4) Attempt *any two* of the following:	12
Q5	(Based on Unit 5) Attempt *any two* of the following:	12

Practical Evaluation (50 marks)

A certified copy journal is essential to appear for the practical examination.

1.	Practical Question 1	20
2.	Practical Question 2	20
3.	Journal	5
4.	Viva Voce	5

OR

1.	Practical Question	40
2.	Journal	5
3.	Viva Voce	5

Project Documentation and Viva Voce Evaluation (50 marks)

The documentation should be checked for plagiarism and as per UGC guidelines, should be less than 10%.

1.	Documentation Report (Chapter 1 to 4)	20
2.	Innovation in the topic	10
3.	Documentation/Topic presentation and viva voce	20

Project Implementation and Viva Voce Evaluation (50 marks)

1.	Documentation Report (Chapter 5 to last)	20
2.	Implementation	10
3.	Relevance of the topic	10
4.	Viva Voce	10

Appendix – 1

Project Documentation and Viva-voce (Semester III) and Project Implementation and Viva-Voce (Semester IV)

Goals of the course Project Documentation and Viva-Voce

The student should:

- be able to apply relevant knowledge and abilities, within the main field of study, to a given problem
- within given constraints, even with limited information, independently analyse and discuss complex inquiries/problems and handle larger problems on the advanced level within the main field of study
- reflect on, evaluate and critically review one's own and others' scientific results
- be able to document and present one's own work with strict requirements on structure, format, and language usage
- be able to identify one's need for further knowledge and continuously develop one's own knowledge

To start the project:

- Start thinking early in the programme about suitable projects.
- Read the instructions for the project.
- Attend and listen to other student´s final oral presentations.
- Look at the finished reports.
- Talk to senior master students.
- Attend possible information events (workshops / seminars / conferences etc.) about therelated topics.

Application and approval:

- Read all the detailed information about project.
- Finalise finding a place and supervisor.
- Check with the coordinator about subject/project, place and supervisor.
- Write the project proposal and plan along with the supervisor.
- Fill out the application together with the supervisor.
- Hand over the complete application, proposal and plan to the coordinator.
- Get an acknowledgement and approval from the coordinator to start the project.

During the project:

- Search, gather and read information and literature about the theory.
- Document well the practical work and your results.
- Take part in seminars and the running follow-ups/supervision.
- Think early on about disposition and writing of the final report.
- Discuss your thoughts with the supervisor and others.
- Read the SOP and the rest you need again.

- Plan for and do the mid-term reporting to the coordinator/examiner.
- Do a mid-term report also at the work-place (can be a requirement in some work-places).
- Write the first draft of the final report and rewrite it based on feedback from the supervisor and possibly others.
- Plan for the final presentation of the report.

Finishing the project:

- Finish the report and obtain an OK from the supervisor.
- Ask the supervisor to send the certificate and feedback form to the coordinator.
- Attend the pre-final oral presentation arranged by the Coordinator.
- Rewrite the final report again based on feedback from the opponents and possibly others.
- Prepare a title page and a popular science summary for your report.
- Send the completed final report to the coordinator (via plagiarism software)
- Rewrite the report based on possible feedback from the coordinator.
- Appear for the final exam.

Project Proposal/research plan

- The student should spend the first 1-2 weeks writing a 1-2 pages project plan containing:
- Short background of the project
- Aims of the project
- Short description of methods that will be used
- Estimated time schedule for the project
- The research plan should be handed in to the supervisor and the coordinator.
- Writing the project plan will help you plan your project work and get you started in finding information and understanding of methods needed to perform the project.

Project Documentation

- The documentation should contain:
- Introduction - that should contain a technical and social (when possible) motivation of the project topic.
- Description of the problems/topics.
- Status of the research/knowledge in the field and literature review.
- Description of the methodology/approach. (The actual structure of the chapters here depends on the topic of the documentation.)
- Results - must always contain analyses of results and associated uncertainties.
- Conclusions and proposals for the future work.
- Appendices (when needed).
- Bibliography - references and links.

For the master's documentation, the chapters cannot be dictated, they may vary according to the type of

project. However, in Semester III Project Documentation and Viva Voce must contain at least 4 chapters

(Introduction,Review of Literature, Methodology / Approach, Proposed Design / UI design, etc. depending on the type of project.)

The Semester III report should be spiral bound.

In Semester IV, the remaining Chapters should be included (which should include Experiments performed, Results and discussion, Conclusions and proposals for future work, Appendices) and Bibliography - references and links. Semester IV report should include all the chapters and should be hardbound.