SIES (Nerul) College of Arts, Science and Commerce
(Autonomous)

B.Sc. Data Science

Semester III and IV

(To be implemented from the Academic Year 2025-26)

AC: 22/02/2025
Item No.: 3.4

SIES (Nerul) College of Arts, Science & Commerce (Autonomous)
RISE WITH EDUCATION

SIES (Nerul) College of Arts, Science and Commerce (Autonomous)
Syllabus for Approval
B.Sc. (Data Science)

Sr. No.	Heading	Particulars
1	Title of the course	S.Y.B.Sc.(Data Science)
2	Passing Marks	40%
3	Semesters	III and IV
4	Level	UG
5	Pattern	3-4 years & 6-8 semesters Choice Based Grading System
6	Status	New
7	To be implemented from	From Academic year 2025-26 in a progressive manner

Date: 22nd February, 2025

Signature:

Dr. Koel Roychoudhury
AC Chairperson

SIES (NERUL) COLLEGE OF ARTS, SCIENCE AND COMMERCE (AUTONOMOUS) NERUL NAVI MUMBAI 400 706.

Dr. Nutan Sawant
Course Coordinator

PREAMBLE

Data is the most important asset in this era of digital revolution. The technological innovations are seen in all walks of life and therefore we are flooded with massive data. Every business relies on data to deliver better products as well as services. The study of data science has become essential to meet the growing demand for data scientists and data analysts.

The application of numerous tools and techniques in the fields of computer science, mathematics, and statistics gave rise to the field of data science. The need to gather and evaluate the massive amounts of data found in various application domains is growing.

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

Programme Outcomes

PO1: To utilize your understanding of computers to examine practical applications.

PO2: Capacity to deal with the ever-changing technical environment, recognizes and analyzes large amounts of data, and modify and adapt in order to support the expansion of the IT industry.

PO3: Create models and programs to address the different domain-specific issues.

PO4: Understand the many forms and types of data to be able to solve the intricate difficulties in the field of data science.

Semester III

Course Code	Course Type	Course Title	Credit
U25DS3MJ01	Major	Data Structures and Algorithms using python	3
U25DS3MJP01		Data Structures and Algorithms using python Practical	1
U25DS3MJ02	Major	Foundations of Machine Learning	3
U25DS3MJP02		Foundations of Machine Learning Practical	1
U25DS3MI01	Minor	Hypothesis Testing and Statistical Inference	3
U25DS3MIP01		Hypothesis Testing and Statistical Inference Practical	1
	OE (any one)	Subject selected from another department (OE basket)	2
U25DS3VSC01	VSC	Database Management Systems	1
U25DS3VSCP01		Database Management Systems Practical	1
U24ENG3AEC01(Revised 2025-2026)	AEC (any one)	Understanding Basic Forms of English Literature-1	2
		Hindi	
		Marathi	
U25CC3NSS03	CC (any one)	National Service Scheme (NSS) Studies Paper-III	4
U25CC3DLLE03		DLLE - Social Work Performance	
U25CC3SP03		Sports-Psychology in Sports and Adapted Physical education	
U25CC3DC01		SIESITDevClub-Learning	
	Total		**22**

Open Elective (To be Opted)

Sr. No.	Course Code	Course Name	Credits
1	U24MMC3E01	Social Media Marketing(BAMMC)	2
2	U24ES3E01	Natural Resource Management(EVS)	2
3	U24BE3E01	Introduction to the Indian Economy(Economics)	2
4	U24MS3E01	Personality Development II(BMS)	2
5	U24CS3E01	Multimedia & Designing(CS)	2
6	U24BI3E01	Basics of Insurance(BI)	2

Semester IV

Course Code	Course Type	Course Title	Credit
U25DS4MJ01	Major	Data Analytics and visualization	3
U25DS4MJP01		Data Analytics and Visualization Practical	1
U25DS4MJ02	Major	Advanced Machine Learning Techniques	3
U25DS4MJP02		Advanced Machine Learning Techniques Practical	1
U25DS4MI01	Minor	Linear Algebra	3
U25DS4MIP01		Linear Algebra Practical	1
	OE (any one)	Subject selected from another department (OE basket)	2
U25DS4SEC01	SEC	Introduction to Data Warehousing	1
U25DS4SECP01		Introduction to Data Warehousing Practical	1
U24ENG4AEC01(Revised 2025-2026)	AEC (any one)	Understanding Basic Forms of English Literature-2	2
		Hindi	
		Marathi	
U25CC4CEP01	CEP	Community Engagement Program	4
	Total		**22**

Open Elective (To be Opted)

Sr. No.	Course Code	Course Name	Credits
1	U24BE4E01	Introduction to International Economics (Economics)	2
2	U24MMC4E01	Photography (BAMMC)	2
3	U24ES4E01	Toxicology and Risk Assessment (EVS)	2
4	U24MS4E01	Digital Marketing Tools (BMS)	2
5	U24COM4E01	Advertising and Brand Management II (Commerce)	2
6	U24BI4E01	Financial Literacy (BI)	2
7	U24IT4E01	Marketing Analytics Using Python (IT)	2
8	U24PT4E01	Packaging Design and Development (PT)	2
9	U24CS4E01	Advance Multimedia and Designing (CS)	2
10	U24AF4E01	Investment Management (AF)	2

Semester - III

Major (3+1)

Data Structures and Algorithms Using Python

COURSE CODE: U25DS3MJ01 **COURSE CREDIT: 03**

1 credit - 15 lectures **1 lecture is 60 minutes**

Course Objectives:

This course aims to:

1. Understand various data structures and their complexities, including time and space analysis.
2. Implement and analyze linear data structures, such as arrays, linked lists, stacks, and queues.
3. Implement and analyze non-linear data structures, including trees and graphs. searching and sorting algorithms.

Course Outcome:

By the end of this course, students will be able to:

1. Analyze various data structures and their complexities, including time and space efficiency.
2. Implement and evaluate linear data structures, such as arrays, linked lists, stacks, and queues, for real-world applications.
3. Implement and apply non-linear data structures, including trees and graphs, for problem-solving. Develop efficient searching and sorting algorithms.

Unit	Topics	Lectures
I	Introduction to Data Structures & Algorithm Analysis: Introduction to Data Structures: Need for Data Structures in Data Science, Types of Data Structures: Linear vs. Non-Linear, Abstract Data Types (ADT) Algorithm Analysis: Time & Space Complexity, Importance of Algorithm Analysis ,Complexity of an Algorithm ,Asymptotic Analysis and Notations. Big O, Big Omega, Big Theta Notation, Rate of Growth, Big O Notation, Recursion & Recursive Algorithms Python Basics for Data Structures: Lists, Tuples, Dictionaries, Sets, Comprehensions & Iterators, Python's Built-in Data Structures and Performance Considerations	15

II	Linear Data Structures & Their Applications: Array and List-Based Structures: Array Representation & Operations, Dynamic Arrays (Python Lists Internals) Linked Lists: Introduction, Representation of Linked List, Linked List vs. Array, Types of Linked List: Singly, Circular, Doubly Linked List, Operations on Singly Linked List and Doubly Linked List, Stack and Queue using Singly Linked List Stack and Queue: Stack: LIFO Principle, Applications (Backtracking, Expression Evaluation, Infix to Postfix Conversion and Postfix Evaluation), Queue: FIFO Principle, Types of Queue, Circular Queue, Priority Queue, Introduction of Double Ended Queue, Applications of Queue Hashing & Hash Tables: Hashing Techniques: Hash Function, Address Calculation Techniques , Common Hashing Functions , Collision Resolution Techniques	15
III	Non-Linear Data Structures & Algorithmic Techniques: Trees: Introduction, Tree Terminologies, Binary Tree, Binary Tree Representation, Types of Binary Tree, Binary Tree Traversals, Binary Search Tree Operations and Applications, Expression Tree, Huffman Encoding, Introduction to AVL tree and B Tree Graph Theory: Introduction, Graph Terminologies, Graph Representation, Graph Traversals: DFS and BFS , Dijkstra's Algorithm, Bellman, Ford Algorithm Sorting & Searching Algorithms: Divide & Conquer: Merge Sort, Quick Sort , Heap Sort, Bubble Sort, Selection Sort, Insertion Sort , Binary Search vs. Linear Search,	15

References:

- Data Structures and Algorithms using Python -Rance D Necaise,John Wiley & Sons
- A Simplified Approach to Data Structures Lalit Goyal, Vishal Goyal, Pawan Kumar SPD 1 st edition ,2014
- Reema Thareja, "Data Structures using C", Oxford Press
- Data Structures and Algorithms in Python" – Michael T. Goodrich, Roberto Tamassia, Michael H. Goldwasser

Data Structures and Algorithms Using Python Practical

COURSE CODE: U25DS3MJP01 **COURSE CREDIT: 01**

1 credit - 15 lectures **1 lecture is 120 minutes**

Course Objectives:

This course aims to:

Implement basic data structures such as linear and non-linear data structures.

Course Outcome:

By the end of this course, students will be able to:

Choose appropriate data structure and apply it in various problems.

Sr. No.	Practical
1	Implement a Python program to perform insert, delete, and search on a list.
2	Write a recursive and iterative Fibonacci function. Compare the performance (time complexity).
3	Implement Stack ADT using array.
4	Convert an Infix expression to Postfix expression using stack ADT.
5	Evaluate Postfix Expression using Stack ADT.
6	Implement Linear Queue ADT using array.
7	Implement Circular Queue ADT using array.
8	Implement Priority Queue ADT using array.
9	Implement Singly Linked List ADT.
10	Implement Circular Linked List ADT.
11	Implement Doubly Linked List ADT.
12	Implement Stack / Linear Queue ADT using Linked List.
13	Implement Binary Search Tree ADT using Linked List.
14	Implement Graph Traversal techniques:) Depth First Search b) Breadth First Search
15	Implement Merge Sort and test its performance.

Major (3+1)

Foundations of Machine Learning

COURSE CODE: U25DS3MJ02 **COURSE CREDIT: 03**

1 credit - 15 lectures **1 lecture is 60 minutes**

Course Objectives:

This course aims to:

- Understand the fundamentals of AI and Machine Learning, including key concepts and applications.
- Analyze the EDA and ML fundamentals
- Understand Linear Regression

Course Outcome:

By the end of this course, students will be able to:

- Explain the relationship between AI and ML. Apply data preprocessing techniques to prepare datasets for machine learning models
- Implement the EDA process.
- Analyze the Linear Regression concept in depth.

Unit	Topics	Lectures
I	**Introduction to AI & Machine Learning:** AI vs. ML vs. Deep Learning, Applications of Machine Learning in Real-World Problems, Types of ML: Supervised, Unsupervised, Reinforcement Learning **Python for Machine Learning:** Introduction to ML Libraries: NumPy, Pandas, Matplotlib, Scikit-Learn, Exploratory Data Analysis (EDA)	15
II	**Exploratory Data Analysis (EDA):** Info, shape, handling missing values, outliers, skewness, Data encoding, feature scaling (Normalization & Standardization), Feature engineering **ML Fundamentals :** ML Modeling Flow, Parametric & Non-parametric Algorithms	15

III	**Linear Regression:** Linear Regression with OLS, Linear Regression with SGD, assumptions, Evaluating Model Parameters, Polynomial Regression, Measuring Performance metrics-Lost and Cost Function (MAE, MSE, RMSE, R2 score), L1 and L2 Regularization, cross-validation **Logistic Regression:** Logistic Regression MLE, evaluating model performance, measuring performance metrics: Confusion Matrix, precision, recall, F1 score, ROC AUC etc., Bias and Variance tradeoff, overfitting and underfitting	15

References:

1. Machine Learning for Humans by Vishal Maini and Samer Sabri
2. Machine Learning For Absolute Beginners by Oliver Theobald
3. "Pattern Recognition and Machine Learning" – Christopher M. Bishop
4. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" – Aurélien Géron

Major (3+1)

Foundations of Machine Learning Practical

COURSE CODE: U25DS3MJP02 **COURSE CREDIT: 01**

1 credit - 15 lectures **1 lecture is 120 minutes**

Course Objectives:

This course aims to:

Equip students with hands-on experience in Machine Learning by implementing real-world data processing.

Course Outcome:

By the end of this course, students will be able to:

Implement the different concept of Machine Learning.

Practical No.	Topic
1	Write a Python program to classify given real-world applications into AI, ML, and Deep Learning
2	Load a dataset and determine whether it is best suited for Supervised, Unsupervised, or Reinforcement Learning.
3	Write a Python script to demonstrate basic operations using NumPy, Pandas, Matplotlib, and Scikit-Learn
4	Load a dataset, display .info(), .shape(), and check for missing values.
5	Fill missing values in a dataset using mean, median, or mode and visualize the effect.
6	Use boxplots to visualize outliers and remove them using the IQR method
7	Convert categorical data into numerical format using One-Hot Encoding and Label Encoding.
8	Apply Min-Max Scaling and Standardization to normalize numerical features.

9	Extract useful features from a dataset, such as date-time components from timestamps.
10	Implement Ordinary Least Squares (OLS) regression using statsmodels and interpret the coefficients.
11	Fit a polynomial regression model and compare its performance with linear regression.
12	Compute MAE, MSE, RMSE, and R^2 score for a trained Linear Regression model.
13	Train a Logistic Regression model and compute the confusion matrix, precision, recall, and F1-score.
14	Plot the ROC curve and calculate the AUC score for a logistic regression model.
15	Compare the performance of an overfitted vs. underfitted model using train-test error analysis.

Minor (3+1)
Hypothesis Testing and Statistical Inference

COURSE CODE: U25DS3MI01 **COURSE CREDIT:3**

1 credit – 15 lectures **1 lecture is 60 minutes**

Course Objectives:

- This course is designed to introduce the concepts of estimation and testing of hypotheses.
- This course also deals with the concept of parametric tests for large and small samples. It also provides knowledge about non-parametric tests and its applications.
- This course will enable students to understand the concept of estimation, test of hypothesis and to apply appropriate estimation techniques and test of hypothesis.

Course Outcomes:

Upon completion of this course, students will be able to

- Demonstrate the concepts of point and interval estimation and use point estimators for estimating unknown parameters.
- Use sampling distributions in testing of hypotheses.
- Apply various parametric and nonparametric tests for one sample and two samples and interpret their results.

UNIT	TOPIC	Lectures
I	**Concept of Population:** Sample Population, Concept of a random sample, parameter, statistic, estimator, unbiased estimator, bias, sampling distribution. standard error of an estimator. Central Limit Theorem for Independently and Identically Distributed Random Variables (Statement only). Sampling Distribution of sample mean, and sample proportion based on large samples. Confidence Interval for population mean and population proportion based on large samples. **Theory of Estimation Point Estimation:** Concept of Estimator and Estimate-properties of Point estimator – consistency, Consistency- Efficiency- relative efficiency- Minimum variance unbiased estimators- Sufficiency- Cramer Rao Inequality (Statement only)- Rao Blackwell Theorem (Statement only)- Neyman Factorization Theorem (Statement only), case study	15

	Tests of Significance I:	
II	**Concept of Statistical hypothesis:**Basics of testing of hypotheses, null and alternative, simple and composite hypotheses. Test of a statistical hypotheses. Critical Region, Probability of Type I and Type II errors, Level of Significance and power of the test, Neyman-Pearson lemma (Statement only). **Large sample tests:** Large sample tests (using central limit theorem, if necessary) For testing specified value of population mean For testing specified value in difference of two means For testing specified value of population proportion For testing specified value of difference of population proportion, Case study concerning Data Science	**15**
III	**Tests of Significance II:** **Sampling distributions of Chi-square- t and F statistics:** derivation of Mean-variance- M.G.F and properties. Small sample tests: Tests for single mean- equality of two means- single variance and equality of two variance- Tests of proportions based on t and F statistics. ANOVA-test for equality of several means. Chi-square tests for independence of attributes and goodness of fit. Case study concerning Data Science	**15**

References:

1. V. K. Rohatgi- Statistical Inference- Dover Publication- New York- 2013.
2. S. C. Gupta and V. K. Kapoor- Fundamentals of Mathematical Statistics-12th ed.- Sultan Chand & Sons- New Delhi- 2017.

Additional References:

1. R. E. Walpole, R. H. Myers and S. L. Myers- Probability and Statistics for Engineers and Scientists- 9th ed.- Pearson- New Delhi- 2017.
2. V. John- Using R for Introductory Statistics- 2nd ed.- CRC Press- Boca Raton- 2014.
3. M. Rajagopalan and P. Dhanavanthan- Statistical Inference-1st ed. - PHI Learning (P) Ltd.-New Delhi- 2012.
4. V. K. Rohatgi and E. Saleh- An Introduction to Probability and Statistics- 3rd ed.- John Wiley & Sons Inc- New Jersey- 2015.

Hypothesis Testing and Statistical Inference Practical

COURSE CODE: U25DS3MIP01 **COURSE CREDIT:1**

1 credit – 15 lectures **1 lecture is 120 minutes**

S.NO.	LIST OF PRACTICALS
1	Practical to obtain Sampling distribution of various statistic.
2	practical to verify unbiasedness and consistency of estimator.
3	Practical to verify efficiency of the estimator.
4	Practical to find the confidence interval for population mean and proportion.
5	Practical to test for single mean and equality of two means when variance is known under normality conditions.
6	Practical to perform one sample t-test.
7	Practical to perform two sample t-test.
8	Practical to perform equality of population variances F-test.
9	Practical to test for single proportion and equality of two proportions.
10.	Practical to test for equality of two proportions.
11	Practical to test for variance One Way ANOVA.
12	Practical to test for independence of attributes using Chi-Square test.
13	Practical to test goodness fit using Chi-Square test.

OE (2)

Subject selected from another department (OE basket)

COURSE CODE:

1 credit - 15 lectures

COURSE CREDIT: 02

1 lecture is 60 minutes

Open Elective (To be Opted)

Sr. No.	Course Code	Course Name	Credits
1	U24MMC3E01	Social Media Marketing(BAMMC)	2
2	U24ES3E01	Natural Resource Management(EVS)	2
3	U24BE3E01	Introduction to the Indian Economy(Economics)	2
4	U24MS3E01	Personality Development II(BMS)	2
5	U24CS3E01	Multimedia & Designing(CS)	2
6	U24BI3E01	Basics of Insurance(BI)	2

VSC (1+1)
Database Management Systems

COURSE CODE: U25DS3VSC01 **COURSE CREDIT: 01**

1 credit - 15 lectures **1 lecture is 60 minutes**

Course Objectives:

This course aims to:

1. To understand data model.
2. To understand database design by normalization.
3. To explore relational database design and dependencies.

Course Outcome:

By the end of this course, students will be able to:

1. Define and describe the fundamental of database architecture.
2. Learners will be able to understand relational database model
3. Learners will be able to do database design using all normalization techniques.
4. Learners can implement various dependencies in the database.

Sr. No.	Topics	Hrs
I	**Data Models:** The importance of data models, Basic building blocks, Business rules,The evolution of data models, Degrees of data abstraction. **Database design and ER Model:** overview, ER Model, Constraints, ER Diagrams, weak entity sets, Codd's rules, Relational Schemas, Relational database model. **Database Design theory and normalization:** Basics of functional dependencies and normalization ((1NF, 2NF, 3NF, BCNF) for relational databases. Relational database design and further dependencies. **Transaction management and Concurrency:** Control Transaction management: ACID properties, serializability and concurrency control, Lock based concurrency control (2PL, Deadlocks), Time stamping methods, optimistic methods, database recovery management.	15

References:

1. "Fundamentals of Database System", Elmasri Ramez, Navathe Shamkant, Pearson Education, Seventh edition, 2017

2. Database Management Systems", Raghu Ramakrishnan and Johannes Gehrke, 3rd Edition, 2014

3. Database Systems: Design implementation and management by Carlos Coronel, Steven Morris, Peter Rob

4. Database System Concepts, Silberschatz, A., Korth, H. F., & Sudarshan, S. (2011), 6th edition. Tata McGraw-Hill Education.

VSC (1+1)

Database Management Systems Practical

COURSE CODE: U25DS3VSCP01
1 credit - 15 lectures

COURSE CREDIT: 01
1 lecture is 120 minutes

Course Objectives:

This course aims to:

- Understand Basic PL/SQL Concepts.
- Write Executable Statements.
- Work with Triggers, Exception Handling.

Course Outcome:

By the end of this course, students will be able to:

- Understand and Implement PL/SQL Variables.
- Write Executable PL/SQL Statements.
- Create Triggers, Handle Exceptions.

Sr. No.	List of Practicals
1	PL/SQL Basics a. Write a PL/SQL program to implement the Use of variables. b. Write executable statement.
2	a.Write a PL/SQL program to Interact with Oracle Server b. Write a PL/SQL program to Create anonymous PL/SQL block
3	Control Structure in PL/SQL. a. Write a PL/SQL program Using while loop b. Write a PL/SQL program to implement Do loop
4	a. Write a PL/SQL program to implement For loop b. Write a PL/SQL program to implement GOTO statement
5	a. Create conditional statement using PL/SQL Using if statement b. Create conditional statement using PL/SQL Using if else statement

6	a. Write a PL/SQL program Using else if ladder b. Write a PL/SQL program Using case expression
7	for Creation of Sequence Write a PL/SQL program in PL/SQL
8	a. Create cursor in PL/SQL using Implicit cursor b. Create cursor in PL/SQL using Explicit Cursor
9	a. Create cursor in PL/SQL using Parameterized cursor b. Create cursor in PL/SQL using Cursor for loop
10	Creation of Procedures in PL/SQL
11	Functions in PL/SQL a. Compute and returns the maximum value b. Compute factorial of given number
12	a. Write a PL/SQL program to Create Row level trigger b. Write a PL/SQL program to Create Statement level trigger
13	a. Write a PL/SQL program to Create instead of trigger b. Write a PL/SQL program using Conditional trigger
14	Handling exceptions a. Creation of user defined exception b. Creation of system defined exception
15	Creation of Package in PL/SQL

Semester - IV


SIES NERUL COLLEGE OF ARTS, SCIENCE AND COMMERCE (AUTONOMOUS)
NERUL
NAVI
MUMBAI
400 706.

Major (3+1)

Data Analytics and Visualization

COURSE CODE: U25DS4MJ01 **COURSE CREDIT: 03**

1 credit - 15 lectures **1 lecture is 60 minutes**

Course Objectives:

This course aims to:

- Introduce data types, EDA, and statistical evaluation methods.
- Teach data processing, analysis, and visualization using Python and Excel.
- Develop skills in data storytelling, dashboarding, and ethical visualization.

Course Outcome:

By the end of this course, students will be able to:

- Students will analyze data, perform EDA, and evaluate models using regression and performance metrics.
- Students will clean, analyze, and visualize data using Pandas, Matplotlib, Seaborn, and Excel dashboards.
- Students will create interactive dashboards using Tableau/Power BI and apply ethical visualization practices.

Unit	Topics	Lectures
I	**Introduction to Data Analysis**: Importance and applications, Types of data: Structured, Unstructured, Semi-structured, Understanding the various levels of data, dealing with categorical, variable, Primary data and Secondary data, **Exploratory Data Analysis (EDA):** Descriptive statistics, distributions, and summary statistics, Correlation and feature selection, Aggregation, pivot tables, and grouping,Confusion Matrix- Area Under Curve- Receiver operating characteristic Curve- Statistical methods for Evaluation- Correlation and Regression.	15

II	**Data analytics and Visualization** **Introduction to Data Visualization :** Importance of visual storytelling, Best practices in visualization **Essential Data Libraries for data analytics**: Pandas, NumPy, SciPy. Plotting and visualization with python: Introduction to Matplotlib, Basic Plotting with Matplotlib, Create Histogram, BarChart, Scatter, Pie chart, Box Plot, violin plot using Matplotlib. Introduction to seaborn Library, Multiple Plots, Regression plot, regplot. **Excel - A Business Intelligence platform** Data cleaning using Excel. Appropriate chart selection for strategy presentation. How to make your data stand out with Pivot Tables and charts. Building interactive dashboards in Excel, Data Preparation, creating Visualization and Interactive dashboards using Excel.	15
III	**Data Storytelling and Applications** **Principles of Data Storytelling:** Communicating insights effectively, Choosing the right visualization for different audiences, Dashboarding and Reporting, **Introduction to Visualization Resources:** Introduction to Tableau, Power BI, Creating interactive dashboards, Power BI **Ethical considerations in data visualization:** Avoiding misleading visualizations, Ensuring transparency and accuracy	15

References:

1. A. Maheshwari- Data Analytics made Accessible-Seattle: Amazon Digital Services- 2015.
2. EMC Education Services- Data Science and Big Data Analytics: Discovering-Analyzing[1]Visualizing and Presenting Data- Wiley- 2015.
3. Ben Fry, "Visualizing Data: Exploring and Explaining Data with the Processing Environment", O'Reilly, 1st Edition, 2008.

Practical of Data Analytics and Visualization

COURSE CODE: U25DS4MJP01 **COURSE CREDIT: 01**

1 credit - 15 lectures **1 lecture is 120 minutes**

Course Objectives:

This course aims to:

To develop hands-on skills in data cleaning, analysis, visualization, and dashboard creation using Python, Excel and Power BI

Course Outcome:

By the end of this course, students will be able to:

Students will be able to clean and analyze data, create meaningful visualizations, build interactive dashboards, and effectively communicate insights.

Sr. No	Practical
1	Given a dataset containing student scores in different subjects, use Pandas and NumPy to calculate the mean, median, and standard deviation for each subject. Additionally, filter students who scored above the class average in Mathematics.
2	Using a dataset of monthly sales for a retail store, create a line chart to visualize sales trends over a year. Add appropriate labels, a title, and a grid for better readability.
3	A dataset contains the salaries of employees in five different departments. Create a box plot and a violin plot to compare the salary distribution across departments. What insights can you draw from the plots?
4	Using the built-in "tips" dataset in Seaborn, create a regression plot (regplot) to analyze the relationship between total bill and tip amount. Does the trend indicate a positive correlation?

5	Given a dataset containing temperature variations for four different cities over a year, create a figure with four subplots—each showing a line graph for one city. Ensure the layout is readable and well-organized.
6	Data Cleaning in Excel – Clean a sales dataset by removing duplicates, handling missing values, and formatting data.
7	Chart Selection for Strategy Presentation – Choose and create appropriate charts (bar, line, pie) to present sales trends effectively.
8	Enhancing Data with Pivot Tables and Charts – Use Pivot Tables and Pivot Charts to summarize and analyze regional sales data dynamically.
9	Building Interactive Dashboards in Excel – Create an interactive sales dashboard with Pivot Tables, slicers, and dynamic charts.
10	Data Preparation and Visualization for Reporting – Prepare a dataset, apply conditional formatting, and visualize sales performance trends.
11	Data Import and Transformation – Load a dataset into Power BI, clean missing values, remove duplicates, and format columns using Power Query.
12	Creating Basic Visualizations – Use bar charts, line charts, and pie charts to visualize sales trends, product performance, and regional distribution.
13	Building Interactive Reports – Create a report with slicers, filters, and drill-through options to analyze sales data dynamically.
14	Creating a Power BI Dashboard – Combine multiple visuals, KPIs, and interactive elements to design a sales performance dashboard.
15	Data Modeling and DAX Calculations – Use relationships and DAX functions to calculate total revenue, profit margins, and year-over-year growth.

Major (3+1)

Advanced Machine Learning Techniques

COURSE CODE: U25DS4MJ02 **COURSE CREDIT: 03**

1 credit - 15 lectures **1 lecture is 60 minutes**

Course Objectives:

This course aims to:

1. Learn and implement supervised learning techniques along with model evaluation.
2. Explore unsupervised learning and optimize models using performance metrics

Course Outcome:

By the end of this course, students will be able to:.

1. Implement supervised learning algorithms for predictive modeling.
2. Use unsupervised learning techniques for pattern recognition and clustering.

Unit	Topics	Lectures
I	**Classification Algorithms** **Decision Tree:** Introduction to Decision tree, Entropy and information gain,Gini index,CART and CHAID,performance metrics,pruning techniques **Naïve Bayes:** Introduction, Bayes Theorem, Types of Naïve Bayes, Naïve Bayes Classifier **Random Forest:** Bootstrap Sampling, Bagging (bootstrap aggregation), Introduction to random forest, why random forest? , performance metrics	15

II	**Ensemble Learning & SVM** **Ensemble techniques:** Boosting,AdaBoost,Gradient Boosting,XGBoost **Neighbours(K-NN)** What is KNN? ,KNN algorithm,working of KNN,how to choose value of K **Support vector Machine(SVM):** Understanding of vectors,decision boundary, what are support vectors & hyperplace, what is support vector machine? Working of SVM, kernels and types of kernel, hard marginand soft margin, SVM for multi-class classification	**15**
III	**Unsupervised Learning & Dimensionality Reduction** **PCA:** Principal component analysis, introduction to dimensionality reduction, what is PCA?, computing components in PCA, dimensionality reduction using PCA **K-means:** Introduction to clustering, what is K-means clustering? K-means clustering algorithm, choosing the optimum K value(Elbow method), various distance measures. **Hierarchical clustering:** Introduction to Hierarchical clustering, Dendragram, Types of Hierarchical Clustering: Aaggomerative and Divisive, cluster linkage	**15**

References:

5. Machine Learning for Humans by Vishal Maini and Samer Sabri
6. Machine Learning For Absolute Beginners by Oliver Theobald
7. "Pattern Recognition and Machine Learning" – Christopher M. Bishop
8. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" – Aurélien Géron

Advanced Machine Learning Techniques Practical

COURSE CODE: U25DS4MJP02 | **COURSE CREDIT: 01**

1 credit - 15 lectures | **1 lecture is 120 minutes**

Course Objectives:

This course aims to:

Ensures a progressive learning approach supervised learning, unsupervised learning, and model optimization.

Course Outcome:

By the end of this course, students will be able to:

Implement supervised learning, unsupervised learning, and model optimization.

No.	Practical Question
1	Implement a Decision Tree classifier on the Iris dataset and visualize the tree structure.
2	Calculate entropy and information gain for a small dataset manually and verify with Scikit-Learn.
3	Compare Gini Index and Entropy as splitting criteria in Decision Trees.
4	Implement a Naïve Bayes classifier for spam email classification.
5	Train a Random Forest model on a dataset and compare its accuracy with a Decision Tree.
6	Demonstrate bootstrap sampling using NumPy and explain its role in Bagging.
7	Compare the performance of Decision Tree, Random Forest, and AdaBoost on a dataset.
8	Train a Gradient Boosting and an XGBoost model and compare their performance.
9	Implement k-NN on a dataset and experiment with different values of K using cross-validation.
10	Train an SVM model with linear, polynomial, and RBF kernels and compare their decision boundaries.
11	Apply PCA on a high-dimensional dataset and visualize the variance explained by each principal component.
12	Perform K-Means clustering on a dataset and use the Elbow Method to determine the optimal K.
13	Compare clustering results using Euclidean, Manhattan, and Cosine distance metrics.
14	Implement Agglomerative Hierarchical Clustering and plot a dendrogram.
15	Compare different linkage methods (single, complete, average) in hierarchical clustering.

Minor (3+1)

Linear Algebra

COURSE CODE: U25DS4MI01 **COURSE CREDIT: 03**

1 credit - 15 lectures **1 lecture is 60 minutes**

Course Objectives:

This course aims to:

- To offer the learner the relevant Linear Algebra concepts through Computer Science applications.
- To interpret existence and analyze the solution set of a system of linear equations.
- To formulate, solve, apply, and interpret properties of linear systems.
- To learn about the concept of linear independence of vectors over a field, and the dimension of a vector space.
- To interpret basic concepts of linear transformations, dimension, matrix representation of a linear transformation, and the change of coordinate matrix.

Course Outcome:

By the end of this course, students will be able to:

- Appreciate the relevance and applications of Linear Algebra in the field of Computer Science.
- Understand the concepts through program implementation. Instil a computational thinking while learning linear algebra.
- Express clear understanding of the concept of a solution to a system of equations. Find eigenvalues and corresponding eigenvectors for a square matrix.
- Understand the introduction and application of concepts like vector spaces, inner product spaces, and linear transformations.
- To apply the knowledge of linear algebra to deal with data to solve real-world problems.

Unit NO	Syllabus	No. of Lectures
1	**Vectors:** Vectors are functions, Vector addition, Scalar-vector multiplication, combining vector addition and scalar multiplication, Dictionary-based representations of vectors, Dot-product, Solving a triangular system of linear equations. **The Vector Space:** Definition of vector space, properties, Subspaces-criterion for a subset to be a subspace Linear combination, Span, The geometry of sets of vectors, Vector spaces, Linear systems, homogeneous and otherwise.	15
2	**Matrix:** Matrices as vectors, Column space and row space, Matrix-vector and vector-matrix multiplication in terms of linear combinations, Matrix-vector multiplication in terms of dot-products, Null space, Computing sparse matrix-vector product, Linear functions, 'Matrix-matrix multiplication, Inner product and outer product, From function inverse to matrix inverse. Linear Transformation and its properties. Basis: Coordinate systems, two greedy algorithms for finding a set of generators, Linear dependence, Basis, Unique representation, Change of basis, first look, Computational problems involving finding a basis, range, and kernel- rank and nullity- Rank-Nullity theorem Gaussian elimination: Echelon form, Gaussian elimination over GF(2), Solving a matrix-vector equation using Gaussian elimination	15

3	**Inner Product:** The inner product for vectors over the reals, Orthogonality. **Orthogonalization**: Projection orthogonal to multiple vectors, projecting orthogonal to mutually orthogonal vectors, Building an orthogonal set of generators, orthogonal complement. **Linear Algebra Application to Data Science** - Loss functions - Regularization-covariance Matrix Support Vector Machine Classification. Linear Algebra in dimensionality Reduction - Principal Component Analysis (PCA) - Singular Value Decomposition (SVD)	**15**

REFERENCES

1. M.P. Deisenroth, A. Aldo Faisal and C.H. Ong- Mathematics for Machine Learning- 1st ed. Cambridge University Press- 2020.

2. G. Strang- Linear Algebra and Learning from Data. – 1st ed.- Wellesley-Cambridge Press- 2019.

ADDITIONAL REFERENCES

1. David C. Lay- Linear Algebra and its Applications- 5th ed.-Indian Reprint- Pearson Education Asia- 2018.

2. K. P. Murthy- Machine Learning- a Probabilistic Perspective-1st ed.- MIT Press- 2012.

3. S. H. Friedberg- A. Insel- and L. Spence- Linear algebra- 4th ed.- Pearson- 2015.

4. Gilbert Strang- Linear Algebra and its Applications- 4th ed.- Thomson Brooks/Cole- 2007.

Linear Algebra Practical

COURSE CODE: U25DS4MIP01 **COURSE CREDIT: 01**

1 credit - 15 lectures **1 lecture is 120 minutes**

Course Objectives:

This course aims to:

To understand and apply fundamental concepts of **vectors, matrices, and dimensionality reduction** in computational data analysis using programming.

Course Outcome:

By the end of this course, students will be able to:

Students will be able to implement vector operations, matrix transformations, and machine learning techniques like Support Vector Machines (SVM) and Principal Component Analysis (PCA), which are essential for financial data analysis and other real-world applications.

Sr. No	List of Practical
1	Write a program which demonstrates the following: · Plotting a set of complex numbers · Creating a new plot by rotating the given number by a degree 90, 180, 270 degrees and also by scaling by a number a = 1/2, a = 1/3, a = 2 etc.
2	Write a program to do the following: · Enter a vector u as a n-list · Enter another vector v as a n-list · Find the vector au + bv for different values of a and b
3	Write a program to find the dot product of u and v
4	Vector Applications:Classify given data using support vector machines (SVM)
5	Write a program to check whether the vectors are linearly independent or dependent.
6	Write a program to convert a matrix into its row echelon form. (Order 2).

7	Write a program to do the following: · Find the vector –matrix multiplication of a r by c matrix M with an c-vector u. · Find the matrix-matrix product of M with a c by p matrix N.
8	Basic Matrix Application – I Representation of Image in Matrix Format and Image Transformations
9	Basic Matrix Application – II Perform Image addition and subtraction
10	Write a program to enter a vector b and find the projection of b orthogonal to a given vector u.
11	Write a program to convert a matrix into its reduced row echelon form.
12	Write a program for dimensionality reduction using PCA
13	Write a program for dimensionality reduction using SVD

OE (2)

~~Subject selected from another department (OE basket)~~

COURSE CODE:
1 credit - 15 lectures

COURSE CREDIT: 02
1 lecture is 60 minutes

Open Elective (To be Opted)

Sr. No.	Course Code	Course Name	Credits
1	U24BE4E01	Introduction to International Economics (Economics)	2
2	U24MMC4E01	Photography (BAMMC)	2
3	U24ES4E01	Toxicology and Risk Assessment (EVS)	2
4	U24MS4E01	Digital Marketing Tools (BMS)	2
5	U24COM4E01	Advertising and Brand Management II (Commerce)	2
6	U24BI4E01	Financial Literacy (BI)	2
7	U24IT4E01	Marketing Analytics Using Python (IT)	2
8	U24PT4E01	Packaging Design and Development (PT)	2
9	U24CS4E01	Advance Multimedia and Designing (CS)	2
10	U24AF4E01	Investment Management (AF)	2

SEC (1+1)

Introduction to Data Warehousing

Course Code: U25DS4SEC01 **Course Credit: 01**

1 Credit - 15 Lectures **1 Lecture Is 60 Minutes**

Course Objective:

This course aims to:

1. Learn the basics of Data Warehousing.
2. Explore the main components and structure of data warehouses.
3. Understand the ETL (Extract, Transform, Load) process.
4. Learn how to use data warehouses for analysis and reporting.

Course Outcome:

Upon successful completion of this course, learners will be able to:

1. Understand the fundamentals of Data Warehousing.
2. Identify key components of a Data Warehouse.
3. Analyze the role of OLAP and OLTP.
4. Design and Model a Data Warehouse.

Sr. No.	Topics	Hrs
I	**Introduction to Data Warehousing:** What is a Data Warehouse (DW)? , The role of data warehousing in business intelligence, data warehouse vs. transactional systems, and data lake. **Key Components of a Data Warehouse:** Data sources, ETL (Extract, Transform, Load) processes, Data Staging, and Data Mart, Data Warehouse Architecture: Three-tier and two-tier models. **Applications of Data Warehousing:** OLAP (Online Analytical Processing) and OLTP (Online Transaction Processing), Business intelligence and reporting. **Data Warehouse Design and Modeling:** Dimensional Modeling: Star Schema, Snowflake Schema, and Fact Constellation, Fact Tables and Dimension Tables, Grain of the fact table, Slowly Changing Dimensions (SCD) Types, ER Modeling vs Dimensional Modeling. **ETL (Extract, Transform, Load) Processes:** ETL Concepts: ETL architecture and flow, Data Extraction from various sources, Data	15

	Transformation: Cleaning, Aggregation, and Sorting, Data Loading: Into Data Warehouses and Data Marts, Overview of ETL tools.	

REFERENCES:

1. Paulraj Ponnian, "Data Warehousing Fundamentals", John Wiley.
2. R. Kimpall, "The Data Warehouse Toolkit", John Wiley.
3. W.H. Inmon, "Building the Data Warehouses", Wiley Dreamtech.
4. "Data Warehousing in the Age of Big Data" by Krish Krishnan
5. "The Data Warehouse Lifecycle Toolkit" by Ralph Kimball, Margy Ross, Warren Thornthwaite, and Bob Becker
6. "Data Warehousing: Concepts, Techniques, Products and Applications" by Arun K. Pujari

Introduction to Data Warehousing Practical

~~COURSE CODE: U25DS4SECP01~~ ~~COURSE CREDIT: 01~~

1 credit - 15 lectures **1 lecture is 120 minutes**

Course Objectives:

This course aims to:

1. Learn to install and configure Oracle 11g Release 2, create and manage databases, and set up Listener and repository.
2. Develop and execute OLAP and OLTP queries, create Snowflake schemas, and populate fact and dimension tables.

Course Outcome:

By the end of this course, students will be able to:

1. Develop and execute OLAP and OLTP queries.
2. Create and manage fact and dimension tables, and implement Snowflake schemas.
3. Import, analyze, and visualize data using Excel and BI tools.

Sr. No.	List of Practicals
1	Installation of Oracle 11g Release 2.
2	Show the configuration of Listener.
3	Show the creation of Database.
4	Show the Workspace and Repository Configuration.
5	Create sample OLAP queries for a simple database.
6	Create sample OLTP queries for a simple database.

7	Write SQL queries to populate the Fact Table and Dimension Tables with sample data.
8	Write SQL queries to create the Snowflake schema and load sample data.
9	Write SQL queries to load data into the fact tables and dimension tables.
10	Import the data from (SQL SERVER) & create Pivot Chart.
11	Import the data from (SQL SERVER) & create Pivot Table.
12	Create table in Excel & Create Pivot Chart.
13	Create table in Excel & Create Pivot Table.
14	Use a BI tool like Tableau or Power BI to create visualizations based on the OLAP queries.
15	Mini Project.

Revised Scheme of Examination

Faculty of Science

(Undergraduate Programme)

SCHEME OF EXAMINATION (for 100 marks 3 credits)

The scheme of examination shall be divided into two parts:

- **Internal assessment 40% i.e.40 marks**
- **Semester end examination 60% i.e.60 marks**

(A)Internal Assessment 40 marks

Description	Marks
An internal test of 20 marks Q.1 a. Multiple choice Questions - 05 Marks b. True/False - 05 Marks Q.2. Attempt 2 questions out of 3 questions (5 marks each) - 10 Marks **OR** Online MCQ Test	20
Presentation/Case Studies/Assignments	15
Attendance and Class Participation	5
Total	40

B) Semester End examination 60 marks

PAPER PATTERN

Duration: 2 hours	
Total Marks:60	
Q.1 15 marks OR 15 marks (7 and 8 marks)-Unit 1	15
Q.2 15 marks OR 15 marks (7 and 8 marks)-Unit 2	15
Q.3 15 marks OR 15 marks (7 and 8 marks)-Unit 3	15
Q.4 15 marks-attempt any 3 out of 6 (from Unit 1, Unit 2, Unit 3)	15
Total	60

Passing criteria: Minimum 40% in Internal (16 out of 40) and 40% (24 out of 60) in semester-end examination.

SCHEME OF EXAMINATION (for 50 marks, 2 credits)

The scheme of examination shall be divided into two parts:

- Internal assessment 40% i.e.20 marks
- Semester end examination 60% i.e.30 marks

(A)Internal Assessment 20 marks(Theory)

Description	**Marks**
An internal test of 10 marks Q.1 Multiple choice Questions/True or False - 05 Marks Q.2. Attempt 1 questions out of 2 questions (5 marks each) - 05 Marks **OR** Online MCQ Test	10
Presentation/Case Studies/Assignments	05
Attendance and Class Participation	05
Total	20

OR

(A)Internal Assessment 20 marks(Practical)

Description	**Marks**
Practical Question	10
Journal	05
Viva	05
Total	20

B) Semester End examination 30 marks

PAPER PATTERN

Duration: 1 hour	
Total Marks:30	
Q.1 15 marks OR 15 marks (7 and 8 marks)-Unit 1	15
Q.2 15 marks OR 15 marks (7 and 8 marks)-Unit 2	15
Total	30

OR

PAPER PATTERN (1 credit Theory)

Duration: 1 hour	
Total Marks:30	
Q.1 15 marks OR 15 marks (7 and 8 marks)-Unit 1	15
Q.2 15 marks OR 15 marks (7 and 8 marks)-Unit 1	15
Total	30

Passing criteria: Minimum 40% in Internal (08 out of 20) and 40% (12 out of 30) in semester-end examination.

SCHEME OF PRACTICAL EXAMINATION

(for 50 marks , 1 credit)

Description	Marks
Practical Question 1	20
Practical Question 2	20
Viva	05
Journal	05
Total	50

Passing Standards

The learners to pass a course shall have to obtain a minimum of 40% marks in each head of passing, consisting of Internal Assessment and Semester End Examination. The learners shall obtain a minimum of 40% marks (i.e. 16 out of 40 or 8 Out of 20) in the Internal Assessment and 40% marks in the Semester End Examination (i.e. 24 Out of 60 or 12 Out of 30) separately, to pass the course and a minimum of Grade D, wherever applicable, to pass a particular semester. A learner will be said to have passed the course if the learner passes the Internal Assessment and Semester End Examination together.