PUNYASHLOK AHILYADEVI HOLKAR SOLAPUR UNIVERSITY, SOLAPUR

Name of the Faculty: Science & Technology

Syllabus

B.Sc. Data Science Part-II
(Semester-III and IV)

As per NEP-2020

To be implemented from Academic Year 2026-27

Semester-wise Structure for

B. Sc. Data Science (Honors/Research) Programme

as per NEP-2020

(w.e.f. – June 2026)

Course Type	Course Code	Course Title	Credits	Teaching hours/week		
				T	P	Total
B.Sc. Part-II (Semester-III) Data Science						
Major	**DSC1-3**	Statistical Computing-I	**2**	**2**	**--**	**2**
	DSC1-3 (P)	Data Science Practical (Major) -III	**1**	**--**	**2**	**2**
	DSC1-4	Classification and Regression-I	**2**	**2**	**--**	**2**
	DSC1-4 (P)	Data Science Practical (Major) -IV	**1**	**--**	**2**	**2**
Minor	**DSC2-3**	Numerical Analysis-I	**2**	**2**	**--**	**2**
	DSC2-3 (P)	Data Science Practical (Minor) -I	**1**	**--**	**2**	**2**
	DSC2-4	Statistical Data Analysis-I	**2**	**2**	**--**	**2**
	DSC2-4 (P)	Data Science Practical (Minor) -II	**1**	**--**	**2**	**2**
OE	**OE-3**	Applications of Data Science	**2**	**2**	**--**	**2**
VSC	**VSC1**	Practical related to Major DSC1-3	**2**	**--**	**4**	**4**
	VSC2	Practical Related to Major DSC1-4	**2**	**--**	**4**	**4**
B.Sc. Part-II (Semester-IV) Statistics						
Major	**DSC1-5**	Statistical Computing-II	**2**	**2**	**--**	**2**
	DSC1-5 (P)	Data Science Practical (Major)-V	**1**	**--**	**2**	**2**
	DSC1-6	Classification and Regression-II	**2**	**2**	**--**	**2**
	DSC1-6 (P)	Data Science Practical (Major)-VI	**1**	**--**	**2**	**2**
Minor	**DSC2-5**	Numerical Analysis-II	**2**	**2**	**--**	**2**
	DSC2-5 (P)	Data Science Practical (Minor) -III	**1**	**--**	**2**	**2**
	DSC2-6	Statistical Data Analysis-II	**2**	**2**	**--**	**2**
	DSC2-6 (P)	Data Science Practical (Minor) –IV	**1**	**--**	**2**	**2**
OE	**OE-4**	AI and Data in Everyday Life	**2**	**2**	**--**	**2**
VSC	**VSC3**	Practical related to Major DSC1-5	**2**	**--**	**4**	**4**
	VSC4	Practical Related to Major DSC1-6	**2**	**--**	**4**	**4**

B. Sc. Part-II (Data Science) Semester-III

DSC1-3	Theory	Statistical Computing-I	Credits: 02 Hours: 30

Course Objectives:

1. To introduce fundamental numerical methods for solving nonlinear equations and optimization problems.
2. To develop understanding of root-finding techniques such as Bisection, Regula-Falsi and Newton–Raphson methods.
3. To provide knowledge of numerical integration techniques including Trapezoidal and Simpson’s rules.
4. To familiarize students with basic numerical search and optimization algorithms used in computational problems.
5. To enable students to implement numerical algorithms using programming tools for simple applications.

Course Outcomes:

After completion of this course, students will be able to:

1. Apply classical numerical methods to find roots of nonlinear equations.
2. Use numerical integration techniques to approximate single and double integrals.
3. Implement and analyze basic optimization and search algorithms such as gradient search, grid search and direct search.
4. Apply advanced numerical methods like Muller’s method and Aitken’s extrapolation for computational problems.
5. Develop simple programs to solve numerical problems and interpret computational results effectively.

Course Content

Unit-1: Numerical Methods:

a) Newton Raphson method,
b) Bisection method,
c) Regula falsi method,
d) gradient search method.

Numerical integration using Trapezoidal rule and Simpson’s rule for single and double integrals.

Programming exercise on these methods. **(15 L)**

Unit-2: Numerical algorithms such as direct search, grid search, interpolation search, gradient search, Bisection and Newton-Raphson methods, Mullers method, Aitkens extrapolation, Simple applications of the above methods. **(15 L)**

Reference Books

1. Ryan B. and Joiner B. L. (2001). MINITAB Handbook, 4th Ed. Duxbury.

2. Thisted R. A. (1998). Elements of Statistical Computing, Chapman and Hall.

3. Kennedy William J. Jr., James E. Gentle (1980), Statistical Computing, Marcel Dekkar

DSC1-3 (P)	Practical	Data Science Practical (Major) -III	Credits: 02 Hours: 60

Course Objectives:

1. To introduce students to basic numerical methods for solving mathematical problems computationally.
2. To develop programming skills for implementation of numerical algorithms.
3. To understand numerical techniques for root finding, optimization and numerical integration.
4. To compare the efficiency and applicability of different numerical methods.
5. To provide practical exposure to computational approaches used in scientific and data science applications.

Course Outcomes:

After successful completion of this course, students will be able to:

1. Implement numerical methods for solving nonlinear equations and optimization problems.
2. Apply numerical integration techniques for approximating definite integrals.
3. Compare the performance and convergence of different numerical algorithms.
4. Develop programs for numerical and computational methods using suitable programming tools.
5. Analyze and interpret numerical results in scientific and data-driven applications.

Suggestive List of Practicals (Using R software)

1	Implementation of Bisection Method for solving nonlinear equations.
2	Implementation of Regula-Falsi Method for finding roots of equations.
3	Implementation of Newton–Raphson Method for numerical solution of equations.
4	Comparative study of Bisection, Regula-Falsi and Newton–Raphson methods.
5	Numerical solution of optimization problems using Gradient Search Method.
6	Numerical integration using Trapezoidal Rule.
7	Numerical integration using Simpson’s Rule.
8	Implementation of Interpolation Search Algorithm.
9	Implementation of Muller’s Method for solving nonlinear equations.
10	Implementation of Aitken’s Extrapolation Method and its applications.

DSC1-4	Theory	Classification and Regression-I	Credits: 02 Hours: 30

Course Objectives:

1. To introduce basic concepts of classification and regression techniques.
2. To understand simple predictive models used in data science.
3. To develop elementary skills in model fitting and interpretation.
4. To provide hands-on understanding of supervised learning methods.

Course Outcomes (COs)

After successful completion of this course, students will be able to:

1. Understand the basic concepts of supervised learning, classification and regression techniques.
2. Apply simple linear regression models for prediction and interpretation of data.
3. Use basic classification techniques such as k-NN, Naïve Bayes and Decision Trees for simple datasets.
4. Evaluate the performance of regression and classification models using suitable accuracy measures.

Course Content

Unit–1: Introduction to Supervised Learning and Regression

- Introduction to Data Science and Machine Learning
- Types of Machine Learning: Supervised and Unsupervised Learning
- Concept of Classification and Regression
- Variables and datasets: predictor and response variables
- Introduction to Simple Linear Regression
- Regression line and interpretation of coefficients
- Method of Least Squares
- Fitting of simple linear regression model
- Prediction using regression model
- Measures of model accuracy: MAE, MSE and RMSE
- Simple applications of regression in real-life datasets **(15 L)**

Unit–2: Classification Techniques

- Introduction to Classification Problems
- Binary and Multiclass Classification
- Concept of training and testing datasets
- k-Nearest Neighbour (k-NN) Classification Method
- Distance measures: Euclidean distance
- Naïve Bayes Classification: basic idea and applications
- Decision Tree Classification: concept and simple structure
- Confusion Matrix
- Classification Accuracy, Precision and Recall
- Simple applications of classification techniques using datasets **(15 L)**

Reference Books:

1. Dutt S., Chandramouli S. and Das A. K. (2019). *Machine Learning*, Pearson India Education Services Pvt. Ltd.
2. Kaushik S. (2020). *Artificial Intelligence and Machine Learning*, Khanna Book Publishing Company.
3. Kumar U. D. and Pradhan M. (2019). *Machine Learning using Python*, Wiley India Pvt. Ltd.
4. Parihar B. (2016). *Data Mining and Data Warehousing*, Oxford University Press India. Maheshwari A. (2020). *Data Analytics Made Accessible*, Amazon Digital Services LLC.

DSC1-4 (P)	**Practical**	**Data Science Practical (Major) -IV**	**Credits: 02 Hours: 60**

Course Objectives:

1. Understand the basics of datasets and data visualization techniques.
2. Apply simple regression techniques for modeling and prediction.
3. Evaluate prediction accuracy using standard error measures.
4. Understand basic concepts of classification methods in data analysis.
5. Implement machine learning techniques such as k-NN, Naïve Bayes, and Decision Tree classifiers.

Course Outcomes:

After successful completion of this course, students will be able to:

1. Visualize and interpret datasets using suitable graphical methods.
2. Fit and interpret simple linear regression models.
3. Perform prediction using regression analysis techniques.
4. Compute and interpret MAE, MSE, and RMSE for model evaluation.
5. Apply k-NN and Naïve Bayes methods for classification problems.

Suggestive List of Practicals (Using R software)

1. Introduction to datasets and data visualization.
2. Fitting a simple linear regression model.
3. Prediction using regression analysis.
4. Computation of MAE, MSE and RMSE.
5. Classification using k-NN method.
6. Implementation of Naïve Bayes classifier.
7. Construction of simple Decision Tree classifier.
8. Evaluation using confusion matrix and accuracy measures.

DSC2-3	Theory	Numerical Analysis-I	Credits: 02 Hours: 30

Course Objectives:

1. To introduce fundamental numerical methods for solving nonlinear equations and optimization problems.
2. To develop understanding of root-finding techniques such as Bisection, Regula-Falsi and Newton–Raphson methods.
3. To provide knowledge of numerical integration techniques including Trapezoidal and Simpson's rules.
4. To familiarize students with basic numerical search and optimization algorithms used in computational problems.
5. To enable students to implement numerical algorithms using programming tools for simple applications.

Course Outcomes:

After completion of this course, students will be able to:

1. Apply classical numerical methods to find roots of nonlinear equations.
2. Use numerical integration techniques to approximate single and double integrals.
3. Implement and analyze basic optimization and search algorithms such as gradient search, grid search and direct search.
4. Develop simple programs to solve numerical problems and interpret computational results effectively.

Course Content

Unit-1: Numerical Methods:

a) Newton Raphson method,
b) Bisection method,
c) Regula falsi method,

Numerical integration using Trapezoidal rule and Simpson's rule for single and double integrals.

Programming exercise on these methods. **(15 L)**

Unit-2: Numerical algorithms such as direct search, grid search, interpolation search, gradient search, Bisection and Newton-Raphson methods, Simple applications of the above methods.

(15 L)

Reference Books

1. Ryan B. and Joiner B. L. (2001). MINITAB Handbook, 4th Ed. Duxbury.

2. Thisted R. A. (1998). Elements of Statistical Computing, Chapman and Hall.

3. Kennedy William J. Jr., James E. Gentle (1980), Statistical Computing, Marcel Dekkar

DSC2-3 (P)	**Practical**	**Data Science Practical (Minor) -I**	**Credits: 02** **Hours: 60**

Course Objectives:

1. To introduce students to basic numerical methods for solving mathematical problems computationally.
2. To develop programming skills for implementation of numerical algorithms.
3. To understand numerical techniques for root finding, optimization and numerical integration.
4. To compare the efficiency and applicability of different numerical methods.
5. To provide practical exposure to computational approaches used in scientific and data science applications.

Course Outcomes:

After successful completion of this course, students will be able to:

1. Implement numerical methods for solving nonlinear equations and optimization problems.
2. Apply numerical integration techniques for approximating definite integrals.
3. Compare the performance and convergence of different numerical algorithms.
4. Develop programs for numerical and computational methods using suitable programming tools.
5. Analyze and interpret numerical results in scientific and data-driven applications.

Suggestive List of Practicals

1	Implementation of Bisection Method for solving nonlinear equations.
2	Implementation of Regula-Falsi Method for finding roots of equations.
3	Implementation of Newton–Raphson Method for numerical solution of equations.
4	Comparative study of Bisection, Regula-Falsi and Newton–Raphson methods.
5	Numerical single integration using Trapezoidal Rule.
6	Numerical single integration using Simpson’s Rule.
7	Numerical double integration using Trapezoidal Rule.
8	Numerical double integration using Simpson’s Rule.
9	Implementation of Interpolation Search Algorithm.

DSC2-4	Theory	Statistical Data Analysis	Credits: 02 Hours: 30

Course Objectives:

1. To introduce basic concepts of classification and regression techniques.
2. To understand simple predictive models used in data science.
3. To develop elementary skills in model fitting and interpretation.
4. To provide hands-on understanding of supervised learning methods.

Course Outcomes (COs)

After successful completion of this course, students will be able to:

1. Understand the basic concepts of supervised learning, classification and regression techniques.
2. Apply simple linear regression models for prediction and interpretation of data.
3. Use basic classification techniques such as k-NN, Naïve Bayes and Decision Trees for simple datasets.
4. Evaluate the performance of regression and classification models using suitable accuracy measures.

Course Content

Unit–1: Introduction to Supervised Learning and Regression

- Introduction to Data Science and Machine Learning
- Types of Machine Learning: Supervised and Unsupervised Learning
- Concept of Classification and Regression
- Variables and datasets: predictor and response variables
- Introduction to Simple Linear Regression
- Regression line and interpretation of coefficients
- Method of Least Squares
- Fitting of simple linear regression model
- Prediction using regression model
- Measures of model accuracy: MAE, MSE and RMSE
- Simple applications of regression in real-life datasets **(15 L)**

Unit–2: Classification Techniques

- Introduction to Classification Problems
- Binary and Multiclass Classification
- Concept of training and testing datasets
- k-Nearest Neighbour (k-NN) Classification Method
- Distance measures: Euclidean distance
- Decision Tree Classification: concept and simple structure
- Confusion Matrix
- Classification Accuracy, Precision and Recall
- Simple applications of classification techniques using datasets **(15 L)**

Reference Books:

1. Dutt S., Chandramouli S. and Das A. K. (2019). *Machine Learning*, Pearson India Education Services Pvt. Ltd.
2. Kaushik S. (2020). *Artificial Intelligence and Machine Learning*, Khanna Book Publishing Company.
3. Kumar U. D. and Pradhan M. (2019). *Machine Learning using Python*, Wiley India Pvt. Ltd.
4. Parihar B. (2016). *Data Mining and Data Warehousing*, Oxford University Press India. Maheshwari A. (2020). *Data Analytics Made Accessible*, Amazon Digital Services LLC.

DSC2-4 (P)	**Practical**	**Data Science Practical (Minor) -II**	**Credits: 02** **Hours: 60**

Course Objectives:

1. To introduce students to classical numerical methods for solving nonlinear equations.
2. To develop understanding of iterative techniques such as Bisection, Regula-Falsi and Newton–Raphson methods.
3. To provide knowledge of numerical integration techniques for approximating definite and multiple integrals.
4. To familiarize students with basic numerical algorithms used in search, optimization and approximation problems.
5. To enable students to apply computational techniques for solving simple scientific and engineering problems.

Course Outcomes:

After successful completion of this course, students will be able to:

1. Understand and apply numerical methods for solving nonlinear equations.
2. Use numerical integration techniques such as Trapezoidal and Simpson's rules for single and double integrals.
3. Implement and analyze basic search and optimization algorithms including grid search, direct search and gradient search methods.
4. Apply advanced numerical techniques such as Muller's method and Aitken's extrapolation for problem solving.
5. Develop computational solutions for numerical problems using appropriate programming tools and interpret results.

Suggestive List of Practicals

1. Introduction to datasets and data visualization.
2. Fitting a simple linear regression model.
3. Prediction using regression analysis.
4. Computation of MAE, MSE and RMSE.
5. Classification using k-NN method.
6. Implementation of Naïve Bayes classifier.
7. Construction of simple Decision Tree classifier.
8. Evaluation using confusion matrix and accuracy measures.

OE-3	Theory	Applications of Data Science	Credits: 02 Hours: 30

The course aims to:

1. Introduce students to the basic concept of Data Science in an easy and non-technical manner.
2. Create awareness about how data is used in everyday life and decision-making.
3. Familiarize students with simple real-world applications of data science in various sectors.
4. Develop basic understanding of different types of data and data representation.
5. Sensitize students to issues of data privacy, ethics, and responsible use of data.

Course Outcomes (COs)

After completing the course, students will be able to:

1. Understand the basic meaning and scope of Data Science in simple terms.
2. Recognize the role of data in everyday applications such as social media, banking, healthcare, and education.
3. Identify different types of data and basic methods of data representation.
4. Interpret simple data visualizations such as bar charts, pie charts, and line graphs.
5. Develop awareness about data privacy, ethical issues, and responsible use of digital data.

Course Content

Unit I: Basics and Everyday Applications of Data Science

- Meaning of Data Science (simple introduction)
- Data in daily life (social media, banking, shopping, mobile apps)
- Difference between Data Science, Data Analytics, and Artificial Intelligence (basic awareness only)
- How companies use data for decision making
- Real-life applications:
    - Recommendation systems (YouTube, Netflix, Amazon)
    - Online payments and fraud detection (basic idea)
    - Healthcare and education applications **(15 L)**

Unit II: Understanding Data, Visualization and Ethics

- Types of data: numbers, text, images, videos (basic classification)
- Structured vs unstructured data (simple examples)
- Basic idea of data collection (apps, surveys, websites)
- Data visualization: bar chart, pie chart, line chart (interpretation only)
- Importance of patterns and trends in data
- Data privacy and ethics:
    - Data misuse and fake information
    - Privacy concerns in mobile apps and social media
    - Responsible use of data **(15 L)**

Reference books:

1. Dey S., Banerjee D. and Kairi S. (2023). *Data Science for Beginners*, PHI Learning Private Limited.
2. Uma Maheswari B. and Sujatha R. (2021). *Introduction to Data Science: Practical Approach with R and Python*, Wiley India Private Limited.
3. Raju C. (2023). *Data Science: A Beginner's Guide*, Penguin Random House India Private Limited.

VSC1	SEC/VSC	Practical related to Major DSC1-3	Credits: 02 Hours: 60

Course Objectives:

1. To introduce students to basic numerical methods for solving mathematical problems computationally.
2. To develop programming skills for implementation of numerical algorithms.
3. To understand numerical techniques for root finding, optimization and numerical integration.
4. To compare the efficiency and applicability of different numerical methods.
5. To develop coding skills in Python.

Course Outcomes:

After successful completion of this course, students will be able to:

1. Implement numerical methods for solving nonlinear equations and optimization problems.
2. Apply numerical integration techniques for approximating definite integrals.
3. Compare the performance and convergence of different numerical algorithms.
4. Develop programs for numerical and computational methods using suitable programming tools.
5. Analyze and interpret numerical results using Python in scientific and data-driven applications.

Suggestive List of Practicals (Using Python)

1	Implementation of Bisection Method for solving nonlinear equations.
2	Implementation of Regula-Falsi Method for finding roots of equations.
3	Implementation of Newton–Raphson Method for numerical solution of equations.
4	Comparative study of Bisection, Regula-Falsi and Newton–Raphson methods.
5	Numerical solution of optimization problems using Gradient Search Method.
6	Numerical integration using Trapezoidal Rule.
7	Numerical integration using Simpson's Rule.
8	Implementation of Interpolation Search Algorithm.
9	Implementation of Muller's Method for solving nonlinear equations.
10	Implementation of Aitken's Extrapolation Method and its applications.

VSC2	SEC/VSC	Practical related to Major DSC1-4	Credits: 02 Hours: 60

Course Objectives:

1. Understand the basics of datasets and data visualization techniques.
2. Apply simple regression techniques for modeling and prediction.
3. Evaluate prediction accuracy using standard error measures.
4. Understand basic concepts of classification methods in data analysis.
5. Implement machine learning techniques such as k-NN, Naïve Bayes, and Decision Tree classifiers.

Course Outcomes:

After successful completion of this course, students will be able to:

1. Visualize and interpret datasets using suitable graphical methods.
2. Fit and interpret simple linear regression models.
3. Perform prediction using regression analysis techniques.
4. Compute and interpret MAE, MSE, and RMSE for model evaluation.
5. Apply k-NN and Naïve Bayes methods for classification problems.

Suggestive List of Practicals (Using R software)

1. Introduction to datasets and data visualization.
2. Fitting a simple linear regression model.
3. Prediction using regression analysis.
4. Computation of MAE, MSE and RMSE.
5. Classification using k-NN method.
6. Implementation of Naïve Bayes classifier.
7. Construction of simple Decision Tree classifier.
8. Evaluation using confusion matrix and accuracy measures.

B. Sc. Part-II (Data Science) Semester-IV

DSC1-5	**Theory**	**Statistical Computing-II**	**Credits: 02** **Hours: 30**

Course Objectives:

The course aims to:

1. Provide basic understanding of numerical methods used in solving linear algebraic systems.
2. Introduce iterative techniques for solving real-world computational problems.
3. Develop conceptual understanding of random number generation and simulation techniques.
4. Familiarize students with Monte Carlo simulation and its statistical applications.
5. Enable students to interpret simple computational results using basic tools like Python/R/Excel.

Course Outcomes (COs)

After completing this course, students will be able to:

1. Understand basic matrix operations and methods for solving systems of linear equations.
2. Apply direct and iterative methods such as Gauss elimination and Gauss-Seidel method.
3. Explain the concept of random numbers and generate pseudo-random numbers using simple techniques.
4. Demonstrate basic simulation techniques including Monte Carlo methods for statistical problems.
5. Interpret computational results and apply them to simple statistical and probability problems.

Course Content

Unit I: Numerical Linear Algebra and Matrix Computations **(15 L)**

- Review of matrices and systems of linear equations
- Solution of linear equations:
    - Gauss Elimination Method
    - Gauss-Jordan Method
    - LU Decomposition (concept and simple computation)
- Iterative methods:
    - Jacobi Method
    - Gauss-Seidel Method
- Applications in solving statistical problems

Unit II: Random Number Generation and Simulation Techniques **(15 L)**

- Concept of random numbers and pseudo-random numbers

  Random number generation using Linear Congruential Method
- Generation of random variables:
    - Uniform distribution
    - Normal distribution (Box-Muller method – basic idea)

- Monte Carlo simulation:
    - Estimation of mean, variance, and probability
    - Simulation of simple statistical experiments
- Applications in statistics and probability problems

Reference Books:

1. Sastry S. S. (2012). *Introductory Methods of Numerical Analysis*, PHI Learning Private Limited.
2. Gerald C. F. and Wheatley P. O. (2004). *Applied Numerical Analysis*, Pearson Education.
3. Chapra S. C. and Canale R. P. (2015). *Numerical Methods for Engineers*, McGraw Hill Education.
4. Devroye L. (1986). *Non-Uniform Random Variate Generation*, Springer.
5. Rubinstein R. Y. and Kroese D. P. (2017). *Simulation and the Monte Carlo Method*, Wiley.
6. Miller R. W. and Freund J. E. (2005). *Probability and Statistics for Engineers*, Pearson Education.

DSC1-5 (P)	Practical	Data Science Practical (Major)-V	Credits: 02 Hours: 30

Course Objectives:

The practical course aims to:

1. Develop hands-on skills in solving systems of linear equations using numerical methods.
2. Enable students to implement and compare direct and iterative methods such as Gauss elimination, Gauss-Seidel, and Jacobi methods.
3. Introduce students to random number generation techniques and pseudo-random number concepts.
4. Provide practical exposure to Monte Carlo simulation for statistical estimation and probability problems.
5. Build computational thinking skills using Python/R/Excel for solving mathematical and statistical problems.

Course Outcomes (COs)

After completing this practical course, students will be able to:

1. Solve systems of linear equations using Gauss Elimination, Gauss-Jordan, and LU decomposition methods.
2. Apply iterative methods such as Jacobi and Gauss-Seidel for numerical solutions.
3. Generate and analyze pseudo-random numbers using Linear Congruential Method.
4. Simulate random variables and estimate statistical measures using Monte Carlo techniques.
5. Compare and interpret theoretical and simulated results in statistical and probability problems using computational tools.

Suggestive List of Practicals

1. Perform basic matrix operations (addition, subtraction, multiplication).
2. Solve a system of linear equations using Gauss Elimination method.
3. Solve a system of linear equations using Gauss-Jordan method.
4. Perform LU Decomposition (simple 2×2 or 3×3 system).
5. Implement Jacobi iterative method for solving linear equations.
6. Implement Gauss-Seidel iterative method for solving linear equations.
7. Generate random numbers using Linear Congruential Method (LCG).
8. Generate and analyze uniform random numbers.
9. Generate normal random variables using Box-Muller method
10. Plot distribution of generated random numbers (uniform/normal).

DSC1-6	Theory	Classification and Regression–II	Credits: 02 Hours: 30

Course Objectives:

The course aims to:

1. Extend the understanding of regression techniques beyond simple linear regression to multiple regression models.
2. Introduce students to advanced classification methods such as logistic regression and ensemble-based approaches.
3. Develop the ability to interpret relationships between multiple variables in real-world datasets.
4. Strengthen understanding of model evaluation techniques for both regression and classification problems.
5. Provide awareness of practical applications of statistical learning methods using simple computational tools.

Course Outcomes (COs)

After completing this course, students will be able to:

1. Apply multiple linear regression models for prediction and analysis of real-world data.
2. Interpret the role of multiple predictors and categorical variables in regression models.
3. Explain logistic regression and its use in binary classification problems.
4. Demonstrate understanding of advanced classification techniques such as decision trees and ensemble ideas.
5. Evaluate model performance using metrics such as confusion matrix, ROC curve, and cross-validation concepts.

Course Content

Unit I: Multiple Regression and Model Building (15 L)

- Review of simple linear regression
- Multiple linear regression model
- Interpretation of multiple predictors (coefficients)
- Categorical variables in regression (basic idea using dummy variables)
- Model building concept: selection of variables (introductory level)
- Multicollinearity (basic awareness only)
- Applications of multiple regression in real datasets
- Prediction using regression models

Unit II: Advanced Classification and Model Evaluation (15 L)

- Logistic Regression:
    - Concept of probability-based classification

  - Binary logistic regression (basic idea and interpretation)
- Decision Tree (advanced understanding):
  - Splitting criteria (Gini/Entropy – conceptual only)
  - Tree pruning (basic idea)
- Introduction to Ensemble Methods (concept level only):
  - Random Forest (idea only)
- Model evaluation techniques:
  - Confusion matrix (revision)
  - ROC curve and AUC (conceptual understanding)
- Cross-validation (basic idea of training/testing reliability)
- Applications in real-world classification problems

Reference Books:

1. James G., Witten D., Hastie T. and Tibshirani R. (2021). *An Introduction to Statistical Learning: with Applications in Python*, Springer.
2. Montgomery D. C., Peck E. A. and Vining G. G. (2012). *Introduction to Linear Regression Analysis*, Wiley.
3. Hastie T., Tibshirani R. and Friedman J. (2009). *The Elements of Statistical Learning*, Springer.
4. Kuhn M. and Johnson K. (2013). *Applied Predictive Modeling*, Springer.
5. Shalev-Shwartz S. and Ben-David S. (2014). *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press.
6. Gareth J., Daniela W., Trevor H., Robert T. (2017). *An Introduction to Statistical Learning*, Springer.

DSC1-6 (P)	**Practical**	**Data Science Practical (Major)-VI**	**Credits: 02** **Hours: 30**

Course Objectives

The practical course aims to:

1. Provide hands-on experience in building multiple regression and classification models using real datasets.
2. Develop skills in interpreting regression coefficients and understanding model behavior.
3. Enable students to implement logistic regression, decision tree, and random forest models.
4. Familiarize students with model evaluation techniques such as confusion matrix, ROC curve, and cross-validation.
5. Strengthen the ability to apply machine learning techniques using Python/R tools for real-world problems.

Course Outcomes (COs)

After completing this practical course, students will be able to:

1. Implement multiple linear regression models and interpret their outputs effectively.
2. Apply logistic regression for binary classification problems and analyze results.
3. Construct decision tree and random forest models for classification tasks.
4. Evaluate machine learning models using metrics such as accuracy, precision, recall, ROC curve, and cross-validation.
5. Develop basic predictive models for real-world datasets using regression and classification techniques.

Suggestive List of Practicals

1. Perform simple linear regression on a dataset and interpret results.
2. Fit a multiple linear regression model using a real dataset.
3. Interpret coefficients of multiple predictors in regression output.
4. Perform prediction using a multiple regression model.
5. Encode categorical variables using dummy variables and include in regression model.
6. Identify and interpret multicollinearity using correlation matrix (basic level).
7. Perform variable selection (basic forward/backward or trial method).
8. Evaluate regression model using MAE, MSE, RMSE.
9. Build a binary logistic regression model using dataset.
10. Compare logistic regression vs linear regression for classification problems.

DSC2-5	Theory	Numerical Analysis-II	Credits: 02 Hours: 30

Course Objectives:

The course aims to:
1. Provide basic understanding of numerical methods used in solving linear algebraic systems.
2. Introduce iterative techniques for solving real-world computational problems.
3. Develop conceptual understanding of random number generation and simulation techniques.
4. Familiarize students with Monte Carlo simulation and its statistical applications.
5. Enable students to interpret simple computational results using basic tools like Python/R/Excel.

Course Outcomes (COs)

After completing this course, students will be able to:
1. Understand basic matrix operations and methods for solving systems of linear equations.
2. Apply Gauss elimination for solving system of linear equations.
3. Explain the concept of random numbers and generate pseudo-random numbers using simple techniques.
4. Demonstrate basic simulation techniques including Monte Carlo methods for statistical problems.
5. Interpret computational results and apply them to simple statistical and probability problems.

Course Content

Unit I: Numerical Linear Algebra and Matrix Computations (15 L)

- Review of matrices and systems of linear equations
- Solution of linear equations:
    - Gauss Elimination Method
    - Gauss-Jordan Method
    - LU Decomposition (concept and simple computation)
- Iterative method:
    - Jacobi Method
- Applications in solving statistical problems

Unit II: Random Number Generation and Simulation Techniques (15 L)

- Concept of random numbers and pseudo-random numbers

Random number generation using Linear Congruential Method
- Generation of random variables:
    - Uniform distribution
    - Normal distribution (Box-Muller method – basic idea)
- Monte Carlo simulation:

    - Estimation of mean, variance, and probability
    - Simulation of simple statistical experiments
- Applications in statistics and probability problems

Reference Books:

1. Sastry S. S. (2012). *Introductory Methods of Numerical Analysis*, PHI Learning Private Limited.
2. Gerald C. F. and Wheatley P. O. (2004). *Applied Numerical Analysis*, Pearson Education.
3. Chapra S. C. and Canale R. P. (2015). *Numerical Methods for Engineers*, McGraw Hill Education.
4. Devroye L. (1986). *Non-Uniform Random Variate Generation*, Springer.
5. Rubinstein R. Y. and Kroese D. P. (2017). *Simulation and the Monte Carlo Method*, Wiley.
6. Miller R. W. and Freund J. E. (2005). *Probability and Statistics for Engineers*, Pearson Education.

DSC2-5 (P)	**Practical**	**Data Science Practical (Minor)-III**	**Credits: 02** **Hours: 30**

Course Objectives:

The practical course aims to:

1. Develop hands-on skills in solving systems of linear equations using numerical methods.
2. Enable students to implement and compare direct and iterative methods such as Gauss elimination, Gauss-Seidel, and Jacobi methods.
3. Introduce students to random number generation techniques and pseudo-random number concepts.
4. Provide practical exposure to Monte Carlo simulation for statistical estimation and probability problems.
5. Build computational thinking skills using Python/R/Excel for solving mathematical and statistical problems.

Course Outcomes (COs)

After completing this practical course, students will be able to:

1. Solve systems of linear equations using Gauss Elimination, Gauss-Jordan, and LU decomposition methods.
2. Apply iterative methods such as Jacobi and Gauss-Seidel for numerical solutions.
3. Generate and analyze pseudo-random numbers using Linear Congruential Method.
4. Simulate random variables and estimate statistical measures using Monte Carlo techniques.
5. Compare and interpret theoretical and simulated results in statistical and probability problems using computational tools.

Suggestive List of Practicals

1. Perform basic matrix operations (addition, subtraction, multiplication).
2. Solve a system of linear equations using Gauss Elimination method.
3. Solve a system of linear equations using Gauss-Jordan method.
4. Perform LU Decomposition (simple 2×2 or 3×3 system).
5. Implement Jacobi iterative method for solving linear equations.
6. Generate random numbers using Linear Congruential Method (LCG).
7. Generate and analyze uniform random numbers.
8. Generate normal random variables using Box-Muller method
9. Plot distribution of generated random numbers (uniform/normal).
10. Estimate mean using Monte Carlo simulation.
11. Estimate variance using Monte Carlo simulation.
12. Estimate probability using Monte Carlo simulation method.

DSC2-6	Theory	Statistical Data Analysis-II	Credits: 02 Hours: 30

Course Objectives:

The course aims to:

1. Extend the understanding of regression techniques beyond simple linear regression to multiple regression models.
2. Introduce students to advanced classification methods such as logistic regression and ensemble-based approaches.
3. Develop the ability to interpret relationships between multiple variables in real-world datasets.
4. Strengthen understanding of model evaluation techniques for both regression and classification problems.
5. Provide awareness of practical applications of statistical learning methods using simple computational tools.

Course Outcomes (COs)

After completing this course, students will be able to:

1. Apply multiple linear regression models for prediction and analysis of real-world data.
2. Interpret the role of multiple predictors and categorical variables in regression models.
3. Explain logistic regression and its use in binary classification problems.
4. Demonstrate understanding of advanced classification techniques such as decision trees and ensemble ideas.
5. Evaluate model performance using metrics such as confusion matrix, ROC curve, and cross-validation concepts.

Course Content

Unit I: Multiple Regression and Model Building (15 L)

- Review of simple linear regression
- Multiple linear regression model
- Interpretation of multiple predictors (coefficients)
- Categorical variables in regression (basic idea using dummy variables)
- Model building concept: selection of variables (introductory level)
- Multicollinearity (basic awareness only)
- Applications of multiple regression in real datasets
- Prediction using regression models

Unit II: Advanced Classification and Model Evaluation (15 L)

- Logistic Regression:
    - Concept of probability-based classification
    - Binary logistic regression (basic idea and interpretation)

- Decision Tree (advanced understanding):
    - Splitting criteria (Gini/Entropy – conceptual only)
    - Tree pruning (basic idea)
- Introduction to Ensemble Methods (concept level only):
    - Random Forest (idea only)
- Model evaluation techniques:
    - Confusion matrix (revision)
    - ROC curve and AUC (conceptual understanding)
- Cross-validation (basic idea of training/testing reliability)
- Applications in real-world classification problems

Reference Books:

1. James G., Witten D., Hastie T. and Tibshirani R. (2021). *An Introduction to Statistical Learning: with Applications in Python*, Springer.
2. Montgomery D. C., Peck E. A. and Vining G. G. (2012). *Introduction to Linear Regression Analysis*, Wiley.
3. Hastie T., Tibshirani R. and Friedman J. (2009). *The Elements of Statistical Learning*, Springer.
4. Kuhn M. and Johnson K. (2013). *Applied Predictive Modeling*, Springer.
5. Shalev-Shwartz S. and Ben-David S. (2014). *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press.
6. Gareth J., Daniela W., Trevor H., Robert T. (2017). *An Introduction to Statistical Learning*, Springer.

DSC2-6 (P)	**Practical**	**Data Science Practical (Minor) –IV**	**Credits: 02 Hours: 30**

Course Objectives

The practical course aims to:

1. Provide hands-on experience in building multiple regression and classification models using real datasets.
2. Develop skills in interpreting regression coefficients and understanding model behavior.
3. Enable students to implement logistic regression, decision tree, and random forest models.
4. Familiarize students with model evaluation techniques such as confusion matrix, ROC curve, and cross-validation.
5. Strengthen the ability to apply machine learning techniques using Python/R tools for real-world problems.

Course Outcomes (COs)

After completing this practical course, students will be able to:

1. Implement multiple linear regression models and interpret their outputs effectively.
2. Apply logistic regression for binary classification problems and analyze results.
3. Construct decision tree and random forest models for classification tasks.
4. Evaluate machine learning models using metrics such as accuracy, precision, recall, ROC curve, and cross-validation.
5. Develop basic predictive models for real-world datasets using regression and classification techniques.

Suggestive List of Practicals

1. Perform simple linear regression on a dataset and interpret results.
2. Fit a multiple linear regression model using a real dataset.
3. Interpret coefficients of multiple predictors in regression output.
4. Perform prediction using a multiple regression model.
5. Encode categorical variables using dummy variables and include in regression model.
6. Identify and interpret multicollinearity using correlation matrix (basic level).
7. Perform variable selection (basic forward/backward or trial method).
8. Evaluate regression model using MAE, MSE, RMSE.
9. Build a binary logistic regression model using dataset.
10. Compare logistic regression vs linear regression for classification problems.
11. Build a decision tree classifier using a dataset.
12. Visualize decision tree and interpret splitting rules (Gini/Entropy concept).
13. Perform decision tree pruning (limit depth or nodes conceptually).
14. Implement Random Forest classifier (basic model building).
15. Evaluate model using ROC curve and AUC score.

OE4		AI and Data in Everyday Life	Credits: 02 Hours: 30

Course Objectives

After completing this practical course, students will be able to:

1. Understand the basic concepts of Artificial Intelligence, Data Science, and Machine Learning.
2. Identify the role of AI and data-driven technologies in everyday life.
3. Develop awareness about smart technologies and their applications in different sectors.
4. Understand how digital platforms collect, process, and use data.
5. Recognize ethical, privacy, and security concerns related to AI and digital data.
6. Develop responsible digital behavior and awareness about cyber safety.
7. Explore emerging trends and career opportunities in AI and data-related fields.
8.

Course Outcomes

After successful completion of the course, students will be able to:

1. Explain basic concepts and applications of Artificial Intelligence and Data Science.
2. Identify real-life applications of AI in education, healthcare, banking, entertainment, and communication.
3. Differentiate between AI, Machine Learning, and Data Science at a basic level.
4. Understand the importance of digital awareness, privacy, and responsible use of technology.
5. Recognize ethical issues related to AI, social media, and digital platforms.
6. Interpret the impact of modern technologies on society and daily life.
7. Develop informed and responsible participation in the digital world.

Course Content

Unit I: Introduction to Artificial Intelligence and Smart Technologies **(15 L)**

- Meaning of Artificial Intelligence (simple introduction)
- AI vs Data Science vs Machine Learning (basic awareness)
- AI applications in daily life:
    - Mobile assistants (Siri, Google Assistant)
    - Chatbots and customer support
    - Social media recommendations
- Smart technologies:
    - Smart homes
    - Smart cities
    - Digital maps and navigation
- AI in education, healthcare, banking, and entertainment
- Benefits and limitations of AI

Unit II: Digital Awareness, Ethics and Future Trends **(15 L)**

- Digital footprints and online behavior
- How online platforms collect and use data
- Cyber safety and responsible internet usage

- Fake news and misinformation in digital media
- Ethical concerns in AI and data usage
- Privacy and security awareness
- Future technologies:
    - Generative AI
    - Self-driving vehicles
    - Robotics and automation (basic awareness only)
- Career opportunities in AI and Data Science (introductory idea)

Reference Books:

1. Taulli T. (2019). *Artificial Intelligence Basics: A Non-Technical Introduction*, Apress. Tegmark M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*, Penguin Books.
2. O'Neil C. (2016). *Weapons of Math Destruction*, Crown Publishing Group.
3. Goyal D. and Saini A. (2020). *Artificial Intelligence and Data Science*, BPB Publications, India.
4. Kumar S. and Singh U. (2021). *Introduction to Artificial Intelligence*, Pearson India.
5. Satapathy S. C. and Bhateja V. (2019). *Artificial Intelligence and Machine Learning*, Wiley India.
6. Sinha P. K. and Sinha P. (2016). *Computer Fundamentals*, BPB Publications, India.
7. Godse A. P. and Godse D. A. (2017). *Data Communication and Networking*, Technical Publications, Pune.
8. Rajaraman V. (2018). *Fundamentals of Computers*, PHI Learning, India.
9. Balagurusamy E. (2017). *Fundamentals of Computers*, McGraw Hill Education, India.

VSC3	**SEC/VSC**	**Practical related to Major DSC1-5**	**Credits: 02** **Hours: 60**

Course Objectives

After completing this practical course, students will be able to:

1. Understand numerical techniques for solving systems of linear equations.
2. Apply matrix computation methods in statistical and mathematical problems.
3. Develop algorithms for iterative methods such as Jacobi and Gauss-Seidel methods.
4. Understand the concept of random numbers and pseudo-random number generation.
5. Generate random variables from standard probability distributions.
6. Apply Monte Carlo simulation techniques for estimation and probability problems.
7. Use computational tools for simulation, visualization, and analysis of statistical experiments.
8. Interpret numerical and simulation results in practical statistical applications.

Course Outcomes

After successful completion of the course, students will be able to:

1. Perform matrix operations and solve systems of linear equations using direct numerical methods.
2. Apply iterative numerical techniques and analyze their convergence properties.
3. Generate pseudo-random numbers and simulate random variables from different distributions.
4. Implement Monte Carlo simulation methods for estimation of mean, variance, probabilities, and other statistical quantities.
5. Compare theoretical and simulated results for statistical experiments.
6. Solve practical statistical and probabilistic problems using computational approaches.
7. Develop simple computer programs for numerical linear algebra and simulation techniques.
8. Interpret and communicate computational results effectively in statistical applications.

Suggestive List of Practicals

1. Estimate mean using Monte Carlo simulation.
2. Estimate variance using Monte Carlo simulation.
3. Estimate probability using Monte Carlo simulation method.
4. Solve a statistical normal equation system using Gauss Elimination method.
5. Compare solutions obtained by Jacobi and Gauss-Seidel iterative methods.
6. Study convergence of Jacobi and Gauss-Seidel methods for diagonally dominant matrices.
7. Generate random samples from a discrete distribution using uniform random numbers.
8. Simulate coin tossing and estimate probabilities using Monte Carlo method.
9. Simulate dice throwing experiment and compare theoretical and simulated probabilities.
10. Perform Monte Carlo simulation to approximate the value of π.

VSC4	**SEC/VSC**	**Practical related to Major DSC1-6**	**Credits: 02** **Hours: 60**

Course Objectives

After completing this practical course, students will be able to:

1. Understand regression and classification techniques used in predictive analytics.
2. Develop and interpret multiple regression and logistic regression models.
3. Apply decision tree and Random Forest methods for classification problems.
4. Understand model building concepts including variable selection and multicollinearity.
5. Evaluate predictive models using confusion matrix, ROC curve, AUC, and cross-validation techniques.
6. Analyze real-world datasets using statistical and machine learning approaches.
7. Use computational tools for predictive modeling and data analysis.
8. Interpret model outputs and communicate analytical findings effectively.

Course Outcomes

After successful completion of the course, students will be able to:

1. Fit and interpret multiple regression and logistic regression models using datasets.
2. Apply classification techniques such as Decision Tree and Random Forest methods.
3. Evaluate and compare predictive models using suitable performance measures.
4. Interpret confusion matrix, ROC curve, AUC score, and cross-validation results.
5. Handle categorical predictors and understand basic model-building strategies.
6. Analyze prediction probabilities and classification accuracy in practical problems.
7. Apply regression and classification methods to real-world datasets.
8. Develop basic computational skills for predictive analytics and statistical modeling.

Suggestive List of Practicals

1. Build a decision tree classifier using a dataset.
2. Visualize decision tree and interpret splitting rules (Gini/Entropy concept).
3. Perform decision tree pruning (limit depth or nodes conceptually).
4. Implement Random Forest classifier (basic model building).
5. Evaluate model using ROC curve and AUC score.
6. Prepare and interpret a confusion matrix for classification results.
7. Perform k-fold cross-validation for evaluating model performance.
8. Compare performance of Decision Tree and Random Forest classifiers using accuracy measures.
9. Analyze prediction probabilities obtained from logistic regression model.
10. Apply regression and classification techniques on a real-world dataset and interpret findings.

*****