NIRMA UNIVERSITY

Institute:	Institute of Technology
Name of Programme:	B. Tech All (Other than CSE)
Course Code:	4CS107IE25
Course Title:	Data Science
Course Type:	Interdisciplinary Minor-Elective
Year of Introduction:	2025-26

L	T	Practical Component				C
		LPW	PW	W	S	
3	0	2	-	-	-	4

Course Learning Outcomes (CLO):

At the end of the course, the students will be able to –

1. relate statistical and pre-processing methods as the basis of the data science domain (BL2)
2. select appropriate techniques and computing environments for applications under consideration (BL3)
3. apply and evaluate a variety of algorithms on different types of data (BL3)
4. design new solutions to solve problems in diverse domains. (BL6)

Unit	Contents	Teaching Hours (Total 45)
Unit-I	**Introduction to Data Science:** Data Science and its Importance, Applications of Data Science, Data Science and Related Fields, Different Computing Environments	04
Unit-II	**Mathematical Foundation for Data Science:** Independence, Bayes Theorem, Discrete & Continuous Random Variables, Probability Mass and Density Functions, Cumulative Distribution Functions, Mean and Variance of a Random Variable, Discrete & Continuous Distributions, Numerical Summaries of Data, Frequency Distributions and Histograms, Matrices and Related Concepts, Gradients	08
Unit-III	**Data Preprocessing:** Different types of Data, Handling Missing Values, Data Normalization, Dimensionality Reduction	05
Unit-IV	**Regression Algorithms for Data Science:** Simple and Multiple Linear Regression using Gradient Descent and Normal Equation Methods, Polynomial Regression, Nonlinear Regression, Evaluation Measures	08
Unit-V	**Classification Algorithms for Data Science:** K-Nearest Neighbours, Decision Trees, Naive Bayes, Feed Forward Neural Network and Backpropagation, Evaluation Measures	10
Unit-VI	**Clustering Algorithms for Data Science:** K-means, Hierarchical Agglomerative and Divisive Clustering, Evaluating Clustering, OPTICS, DBSCAN	07
Unit-VII	**Introduction to Deep Learning:** Basics of Deep Learning, Applications of Deep Learning in Data Science	03

Self-Study:

The self-study contents will be declared at the commencement of the semester. Around 10% of the questions will be asked from self-study contents.

Suggested Readings/ References:

1. Jiawei Han, Micheline Kamber, and Jian Pei Data Mining: Concepts and Techniques, Morgan Kaufmann
2. C.M. Douglas and G.C. Runger, Applied Statistics and Probability for Engineers, Wiley
3. Tom Mitchell, Machine Learning, TataMcGraw Hill
4. Athem Ealpaydin, Introduction to Machine Learning, Prentice Hall India
5. Andrew Bruce, Peter C. Bruce, and Peter Gedeck, Practical Statistics for Data Scientists, O'Reilly
6. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Introduction to Statistical Learning, Springer
7. Joel Grus, Data Science from Scratch, O'Reilly

Suggested List of Experiments

S. No.	Title	Hours
1	Setup and introduction to Jupyter Notebook and Python libraries for data science.	04
2	Simulate and visualize different probability distributions using Python.	02
3	Handle missing values, normalize data, and apply dimensionality reduction (PCA) on a sample dataset.	02
4	Implement simple and multiple linear regression models using both gradient descent and normal equation methods on a real-world dataset.	04
5	Implement polynomial regression and evaluate its performance.	02
6	Implement k-NN and decision tree classifiers, and evaluate their performance using metrics like accuracy, precision, recall, and F1-score.	04
7	Implement a Naive Bayes classifier and test its performance on a sample dataset	02
8	Build and train a feed-forward neural network using a deep learning framework and understand the backpropagation algorithm.	02
9	Apply k-means and hierarchical clustering algorithms on a sample dataset and evaluate the clustering results.	04
10	Implement a simple deep learning model for a basic application (e.g., image classification or sentiment analysis) using a deep learning framework.	04