Course code	Course Name	Hours/week			Credit	Max. Marks
	Introduction to Data Science	**L**	**T**	**P**	**C**	**100**
		3	0	1	4	
Pre-Requisite	Basics of Python tool					
Evaluation Scheme	**Theory & Practical**				**Hours**	**Marks**
	External (End Semester Exam)				2	50
	Internal (1) Midterm – 10 Marks (2) CIA (Assignment/Certification/Quiz)-15 Marks (3) Attendance -5 Marks (4) Practical - 20 Marks					50

UNIT-I	**FOUNDATIONS OF DATA SCIENCE AND DATA MANIPULATION**	**9**

Introduction to Python and Data Science Libraries, Fundamental Libraries in Data Science, Integrated development environment (IDE). Data operations: Reading, Selecting, Filtering, Manipulating, Sorting, Grouping, Rearranging, Ranking. Plotting: Basic plots, visualizing distributions and detecting outliers, heatmaps.

UNIT-II	**DATA SUMMARIZATION AND STATISTICAL INFERENCE**	**10**

Descriptive Statistics, Data Preparation: Handling Missing Data, Outliers and Duplicates, Data Scaling and Normalization, Handling Imbalanced Data (basic overview). Exploratory Data Analysis: Data Summarization, Data Distribution, Measuring Asymmetry. Sample and Estimated Mean, Variance and Standard Score. Probability, Random Variable, Probability Distribution Functions, Central Limit Theorem. Statistical Inference: Frequency Approach, Variability of Estimates, Hypothesis Testing Using Confidence Intervals, Using p-value.

UNIT-III	**SUPERVISED LEARNING - 1**	**10**

Supervised Learning (Classification): First Step in Supervised Learning, Learning Curves, Training-Validation and Test. Learning Models Generalities, Metrics for Evaluating the Model, Supervised Learning Classification Methods: Logistic Regression, Support Vector Machines, Random Forest and Linear Discriminant Analysis.

UNIT-IV	**SUPERVISED LEARNING - 2**	**8**

Supervised Learning (Regression Analysis): Linear Regression, Simple Linear Regression, Multiple & Polynomial Regression, Sparse model. Metrics for Evaluating the Model.

UNIT-V	**UNSUPERVISED LEARNING TECHNIQUES**	**8**

Introduction to Unsupervised Learning, Clustering Techniques: K-Means, Hierarchical Clustering, DBSCAN, Similarity and Distances Measures, Quality Measures of Clustering, Metrics for Evaluating the Model, Visualization of Clusters, Case Study: Customer Segmentation using Clustering-Grouping News Articles using Text Clustering-Outlier Detection using DBSCAN

Total hours	**45** periods

Course Outcomes: At the end of the course, students will be able to:

COs	Statements	Bloom's Level
CO1	Understand fundamental Python libraries and perform basic data operations and visualizations.	**L2**
CO2	Apply descriptive statistics and perform statistical inference techniques on real world data.	**L3**

CO3	Implement supervised learning algorithms for classification tasks.	**L3**
CO4	Analyze and evaluate regression models for prediction.	**L4**
CO5	Apply unsupervised learning techniques to extract patterns from unlabeled data.	**L3**
TEXT BOOK:		

1.	VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media. ISBN: 9781491912058
2.	McKinney, W. (2018). *Python for data analysis: Data wrangling with pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media. ISBN: 9781491957660
3.	Motwani, B. (2020). *Data analytics using Python*. Wiley India. ISBN: 9788126502952
REFERENCES:	
4.	Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media. ISBN: 9781492032649
5.	Raschka, S., & Mirjalili, V. (2019). *Python machine learning* (3rd ed.). Packt Publishing. ISBN: 9781789955750
ONLINE REFRENCES:	
1	Python for Data Science Handbook by Jake VanderPlas https://jakevdp.github.io/PythonDataScienceHandbook/ Covers NumPy, pandas, matplotlib, scikit-learn – all key Python libraries for data science.
2	W3Schools Python Tutorial https://www.w3schools.com/python/ Good beginner-level tutorial for Python basics and libraries.
3	Kaggle Learn: Python & Data Science Courses https://www.kaggle.com/learn Free micro-courses on Python, pandas, EDA, ML, statistics, and more.

LIST OF EXPERIMENTS:

S. No.	List of Experiments	COs	No. of Hours
1	Python setup & IDE configuration for Data Science	CO1	2
2	Read, select, filter, and manipulate datasets using pandas		2
3	Sort, group, and rank data; create heatmaps using seaborn		2
4	Handle missing data and outliers in datasets	CO2	2
5	Perform data scaling and normalization		2

6	Generate descriptive statistics summary and visualize distributions		2
7	Hypothesis testing using confidence intervals and p-values		
8	Logistic Regression implementation for classification	CO3	2
9	SVM classification with sklearn		2
10	Random Forest classification and evaluating model performance		2
11	Linear Regression with sklearn	CO4	2
12	Polynomial and Multiple Regression on real datasets		2
13	Sparse model and Lasso/Ridge Regression		2

14	K-Means clustering and quality evaluation	CO5	2
15	Case study on clustering technique with evaluation metrics		2
	Total No. of Hours		**30** Hours

COURSE OUTCOMES WITH PROGRAM OUTCOMES:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO 1	3	2	2	2	2	–	–	–	–	–	–	–
CO 2	3	3	2	2	2	–	–	–	–	–	–	–
CO 3	3	2	3	2	3	–	–	–	–	–	–	–
CO 4	3	3	3	3	2	–	–	–	–	–	–	–
CO 5	3	3	3	2	2	–	–	–	–	–	–	–