

Programme	B. Sc. Computer Science				
Course Title	Introduction to Data Science				
Type of Course	Minor				
Semester	II				
Academic Level	100-199				
Course Details	Credit	Lecture per week	Tutorial per week	Practical per week	Total Hours
	4	3	-	2	75
Pre-requisites	1. Python Programming 2. Linear Algebra				
Course Summary	This course provides a comprehensive overview of data science, covering the various types of data and their applications. The students will acquire a deep understanding of exploratory data analysis along with hands-on implementation skills. . The curriculum introduces both supervised and unsupervised and techniques of Machine learning. Additionally, the data pre-processing techniques are introduced Overall, the course provides a comprehensive understanding of the fundamental data science principles, guiding students through the data science process and illustrating practical applications.				

Course Outcomes (CO):

CO	CO Statement	Cognitive Level*	Knowledge Category#	Evaluation Tools used
CO1	Understand the types of data and the applications of data science	U	C	Instructor-created exams / Quiz
CO2	Implement various data analysis and visualisation techniques	Ap	P	Practical Assignments / Lab exercises / Case studies

CO3	Express and gain insights on advanced data preprocessing techniques	U	P	Instructor-created exams / Quiz / Short analytical assignments
CO4	Discuss various machine learning techniques and algorithms	U	P	Instructor-created exams / Quiz / Case-based questions
CO5	Create analysis and prediction models using machine learning methodologies	C	M	Mini Project / Modelling Assignment / Case study / Demonstration
* - Remember (R), Understand (U), Apply (Ap), Analyse (An), Evaluate (E), Create (C) # - Factual Knowledge(F) Conceptual Knowledge (C) Procedural Knowledge (P) Metacognitive Knowledge (M)				

Detailed Syllabus:

Module	Unit	Content	Hrs	Marks (70)
I	Introduction to Data Science		10	15
	1	Introduction to Data: Types of Data –Structured Data, Semi-Structured Data, Unstructured Data and Data Streams, Statistical Data Types - Quantitative Data (Ratio and Interval Scale) and Qualitative Data (Nominal and ordinal)	2	
	2	Basic Methods of Data Analysis- Descriptive Data Analysis, Diagnostic Data Analysis or Exploratory Data Analysis, Inferential Data Analysis and Predictive Analysis.	1	
	3	Inferential Statistics: Statistical Inference, Population and Sample, Statistical Modeling, Probability Distributions– Normal, Uniform	3	
	4	Introduction to Data Science: Big Data and Data Science , Data Science Process	2	
	5	Applications of Data Science , Issues and challenges in Data Science	2	
II	Exploratory Data Analysis		14	20

	6	Exploratory Data Analysis (EDA) : Types of EDA - Univariate non-graphical, Multivariate non- graphical, Univariate graphical, and, Multivariate graphical. Methods of EDA- Descriptive Statistics and Data Visualization	5	
	7	Descriptive Statistics - Measures of Central Tendencies, Dispersion, Skewness and Kurtosis.	5	
	8	Data Visualization - Histograms , Box plots , Quantile-Quantile plots Scatter plots , Heat map, Bubble chart , Bar chart, Distribution plot , Pair plot , Line graph , Pie chart, Area chart	4	
III	Data Preparation for Analysis		6	15
	9	Data Cleaning: Handling Missing and Noisy Data, Removing outliers	2	
	10	Data Integration	1	
	11	Data Transformation: Standardization, Normalization	2	
	12	Data Reduction: Dimensionality Reduction - Principal Component Analysis	1	
1V	Introduction to Machine Learning		15	20
	13	Machine Learning Algorithms : Supervised Learning- Classification, Regression, Unsupervised Learning-Clustering, Dimensionality Reduction , Reinforcement Learning	3	
	14	Test /Train Split, Model Training, Bias and Variance, Overfitting and Underfitting	3	
	15	Evaluation	2	
	16	Linear Regression	1	
	17	k-Nearest Neighbors (k-NN)	1	
	18	k-means Clustering	1	
	19	Naive Bayes	1	
	20	Application of Naive Bayes - Spam Filtering	1	
	21	Singular Value Decomposition	1	
	22	Applications of Supervised, Unsupervised and Reinforcement	1	

		Learning		
V	Hands-on Data Structures: Practical Applications, Case Study and Course Project		30	20
	1	Implementation of the concepts or the algorithms learned [Binary Classification, Linear Regression, k-NN, k-means clustering, Spam Filtering]	15	
	2	Case study: Perform exploratory data analysis on a real world dataset using Python. Using appropriate Python packages parse, clean and visualize the data .	5	
	3	Capstone/Course Project: Perform an end-to-end project of the data science process.		

Mapping of COs with PSOs and POs :

	PSO1	PSO2	PSO3	PSO4	PSO 5	PSO 6	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	-	-	-	-	-	-	3	1	2	3	1	1	1
CO2	-	-	-	-	-	-	3	1	2	3	3	2	3
CO3	-	-	-	-	-	-	3	1	2	3	3	2	3
CO4	-	-	-	-	-	-	3	1	2	3	3	2	3
CO5	-	-	-	-	-	-	3	1	2	3	3	2	3

Correlation Levels:

Level	Correlation
-	Nil

1	Slightly / Low
2	Moderate / Medium
3	Substantial / High

Assessment Rubrics:

- Quiz / Assignment/ Quiz/Discussion/ Seminar
- Midterm Exam
- Programming Assignments (20%)
- Final Exam (70%)

Mapping of COs to Assessment Rubrics :

	Internal Exam	Assignment	Project Evaluation	End Semester Examinations
CO1	✓	✓		✓
CO2	✓	✓		✓
CO3	✓			✓
CO4	✓			✓
CO5		✓	✓	

References

1. O'Neil, Cathy, and Rachel Schutt. *Doing data science: Straight talk from the frontline.* " O'Reilly Media, Inc.", 2013.
2. Han, Jiawei, et al. *Data Mining: Concepts and Techniques.* Netherlands, Elsevier Science, 2011.
3. Shah, Chirag. *A Hands-On Introduction to Data Science.* United Kingdom, Cambridge University Press, 2020.
4. Chopra, Rohan, et al. *Data Science with Python: Combine Python with Machine Learning Principles to Discover Hidden Patterns in Raw Data.* United Kingdom, Packt Publishing, 2019.