DATA SCIENCE AND APPLICATION

(Effective from the Academic Year 2022 - 2023)

VI SEMESTER

Course Code	**21CS644**	CIA Marks	50
Number of Contact Hours/Week (L: T: P: S)	3:0:0:0	SEE Marks	50
Total Hours of Pedagogy	**40L**	Exam Hours	03

CREDITS – 3

COURSE PREREQUISITES:

COURSE OBJECTIVES:

- To learn data collection and preprocessing techniques for data science
- To Understand and practice analytical methods for solving real life problems.
- To study data exploration techniques
- To learn different types of data and its visualization
- To map element of visualization well to perceive information

TEACHING - LEARNING STRATEGY:

Following are some sample strategies that can be incorporate for the Course Delivery

- Chalk and Talk Method/Blended Mode Method
- Power Point Presentation
- Expert Talk/Webinar/Seminar
- Video Streaming/Self-Study/Simulations
- Peer-to-Peer Activities
- Activity/Problem Based Learning
- Case Studies
- MOOC/NPTEL Courses
- Any other innovative initiatives with respect to the Course contents

COURSE CONTENTS

MODULE – I

Introduction to data Science: Brief History of Data Science, Data Science role and skill tracks, What kind of question can Data sceice solve ,Structure of Data Science Team, Data science roles. ***Textbook 3: Chapter 1***	**8 Hours**

MODULE – II

Statistics: Describing a Single Set of Data, Correlation, Simpson's Paradox, Some Other Correlational Caveats, Correlation and Causation **Probability :** Dependence and Independence, Conditional Probability, Bayes's Theorem, Random Variables, Continuous Distributions, The Normal Distribution. ***Textbook 1: Chapter 5, Chapter 6***	**8 Hours**

MODULE – III

Data Analysis : Getting Data ,Working with Data ,KNN, Simple Linear regression, Multiple regression. ***TextBook 1 : Chapter 9, Chapter 10, Chapter 12,Chapter 14,Chapter 15.***	**8 Hours**

MODULE – IV

Visualizing Data: Numpy , Pandas, Matplotlib,Bar, chart,Line chart,Scatter Plot. ***Textbook 2: Chapter 1*** ***TextBook 1:Chapter 3***	**8 Hours**

MODULE – V

Simplifying Visualizations using Seaborn: Introduction, Advantages of Seaborn Controlling Figure Aesthetics: Seaborn Figure Styles, Removing Axes Spines, Contexts; Color Palettes: Categorical Color Palettes, Sequential Color Palettes, Diverging Color Palettes; Interesting Plots in Seaborn: Bar Plots, Kernel Density Estimation, Plotting Bivariate Distributions, Visualizing Pairwise Relationships, Violin Plots ***Text Book 1: Chapter 4***	**8 Hours**

COURSE OUTCOMES

Upon completion of this course, the students will be able to:

CO No.	Course Outcome Description	Bloom's Taxonomy Level
CO1	Apply data preprocessing methods on open access data and generate quality data for analysis	CL3
CO2	Apply and analyze classification and regression data analytical methods for real life Problems.	CL3
CO3	Implement analytical methods using Python/R.	CL3
CO4	Apply different data visualization techniques to understand the data.	CL3
CO5	Analyze the data using suitable method, visualize using the open source tool.	CL3

CO-PO-PSO MAPPING

CO No.	Programme Outcomes (PO)												Programme Specific Outcome (PSO)	
	1	2	3	4	5	6	7	8	9	10	11	12	1	2
CO1	2	2		2	2				2	2	2	2		
CO2	2	2		2	2				2	2	2	2		
CO3	2	2		2	2				2	2	2	2		
CO4	2	2		2	2				2	2	2	2		
CO5	2	2		2	2				2	2	2	2		

3: Substantial (High)	2: Moderate (Medium)	1: Poor (Low)

ASSESSMENT STRATEGY

Assessment will be both CIA and SEE. Students learning will be assessed using Direct and Indirect methods:

Sl. No.	Assessment Description	Weightage (%)	Max. Marks
1	**Continuous Internal Assessment (CIA)**	**100 %**	**50**
	Continuous Internal Evaluation (CIE)	60 %	30
	Assignments	40 %	20
2	**Semester End Examination (SEE)**	**100 %**	**50**

ASSESSMENT DETAILS

Continuous Internal Assessment (CIA) (50%)				**Semester End Exam (SEE) (50%)**
Continuous Internal Evaluation (CIE) (60%)			**Assignment/ Activities (40%)**	
I	**II**	**III**		
Syllabus Coverage			**Syllabus Coverage**	**Syllabus Coverage**
40%	**30%**	**30%**	**100%**	**100%**
MI			MI	MI
MII	MII		MII	MII
	MIII		MIII	MIII
		MIV	MIV	MIV
		MV	MV	MV

Note: For Examinations (both CIE and SEE), the question papers shall contain the questions mapped to the appropriate Bloom's Level. Any COs mapped with higher cognitive Bloom's Level may also be assessed through the assignments.

ASSIGNMENT TYPES WITH WEIGHTAGES

Sl. No.	Assignment Description	Max. Weightage (%)	Max. Marks
1	Written Assignments	25 %	05
2	Quiz	10 %	02
3	Case Studies	25 %	05
4	Seminar/Presentation	15 %	03
5	Peer - to - Peer Learning	10 %	02
6	Activity Based Learning	50 %	10
7	Project Based Learning	50 %	10
8	Field Work + Report	50 %	10
9	Industry Visit + Report	50 %	10
10	NPTEL/MOOC Courses – Registration and Assignment Submissions	50 %	10
	NPTEL Certification	75 %	15
11	Any other Innovative Assignments (CL4 and above)	50 %	10

Note: The assignments mentioned above may be provided appropriately to the students belonging to different bands

SEE QUESTION PAPER PATTERN:

- The question paper will have **TEN** full questions from **FIVE** Modules
- There will be 2 full questions from each module. Every question will carry a maximum of 20 marks.
- Each full question may have a maximum of four sub-questions covering all the topics under a module.
- The students will have to answer FIVE full questions, selecting one full question from each module.

TEXT BOOKS:

1. Data Science from Scratch : Joel Grus, O'Reilly Media Inc., ISBN: 9781491901427
2. Data Visualisation Workshop,Tim Grobmann and Mario Dobler , Packt Publishing
3. Practitioner's Guide to Data Science : Hui Lin and Ming Li

REFERENCE BOOKS:

1. Big data black book, Dream tech publication
2. Getting Started with Business Analytics: Insightful Decision-Making , David Roi Hardoon, GalitShmueli, CRC Press
3. Business Analytics , James R Evans, Pearson
4. Python Data science Handbook, *Jake VanderPlas, Orielly publication*
5. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, Vovost Foster, Fawcett Tom