

VI Semester

DATA SCIENCE AND VISUALIZATION			
Course Code	21CS644	CIE Marks	50
Teaching Hours/Week (L:T:P: S)	3:0:0:0	SEE Marks	50
Total Hours of Pedagogy	40	Total Marks	100
Credits	03	Exam Hours	03
Course Learning Objectives			
<p>CLO 1. To introduce data collection and pre-processing techniques for data science</p> <p>CLO 2. Explore analytical methods for solving real life problems through data exploration techniques</p> <p>CLO 3. Illustrate different types of data and its visualization</p> <p>CLO 4. Find different data visualization techniques and tools</p> <p>CLO 5. Design and map element of visualization well to perceive information</p>			
Teaching-Learning Process (General Instructions)			
<p>These are sample Strategies, which teachers can use to accelerate the attainment of the various course outcomes.</p> <ol style="list-style-type: none"> Lecturer method (L) need not to be only a traditional lecture method, but alternative effective teaching methods could be adopted to attain the outcomes. Use of Video/Animation to explain functioning of various concepts. Encourage collaborative (Group Learning) Learning in the class. Ask at least three HOT (Higher order Thinking) questions in the class, which promotes critical thinking. Adopt Problem Based Learning (PBL), which fosters students' Analytical skills, develop design thinking skills such as the ability to design, evaluate, generalize, and analyze information rather than simply recall it. Introduce Topics in manifold representations. Show the different ways to solve the same problem with different circuits/logic and encourage the students to come up with their own creative ways to solve them. Discuss how every concept can be applied to the real world - and when that's possible, it helps improve the students' understanding. 			
Module-1			
Introduction to Data Science			
<p>Introduction: What is Data Science? Big Data and Data Science hype – and getting past the hype, Why now? – Datafication, Current landscape of perspectives, Skill sets. Needed Statistical Inference: Populations and samples, Statistical modelling, probability distributions, fitting a model.</p>			
Textbook 1: Chapter 1			
Teaching-Learning Process	<ol style="list-style-type: none"> PPT – Recognizing different types of data, Data science process Demonstration of different steps, learning definition and relation with data science 		
Module-2			
Exploratory Data Analysis and the Data Science Process			
<p>Basic tools (plots, graphs and summary statistics) of EDA, Philosophy of EDA, The Data Science Process, Case Study: Real Direct (online realestate firm). Three Basic Machine Learning Algorithms: Linear Regression, k-Nearest Neighbours (k- NN), k-means.</p>			
Textbook 1: Chapter 2, Chapter 3			
Teaching-Learning Process	<ol style="list-style-type: none"> PPT –Plots, Graphs, Summary Statistics Demonstration of Machine Learning Algorithms 		

Module-3	
Feature Generation and Feature Selection	
Extracting Meaning from Data: Motivating application: user (customer) retention. Feature Generation (brainstorming, role of domain expertise, and place for imagination), Feature Selection algorithms. Filters; Wrappers; Decision Trees; Random Forests. Recommendation Systems: Building a User-Facing Data Product, Algorithmic ingredients of a Recommendation Engine, Dimensionality Reduction, Singular Value Decomposition, Principal Component Analysis, Exercise: build your own recommendation system.	
Textbook 1: Chapter 6	
Teaching-Learning Process	<ol style="list-style-type: none"> 1. PPT – Feature generation, selection 2. Demonstration recommendation engine
Module-4	
Data Visualization and Data Exploration	
Introduction: Data Visualization, Importance of Data Visualization, Data Wrangling, Tools and Libraries for Visualization	
Comparison Plots: Line Chart, Bar Chart and Radar Chart; Relation Plots: Scatter Plot, Bubble Plot , Correlogram and Heatmap; Composition Plots: Pie Chart, Stacked Bar Chart, Stacked Area Chart, Venn Diagram; Distribution Plots: Histogram, Density Plot, Box Plot, Violin Plot; Geo Plots: Dot Map, Choropleth Map, Connection Map; What Makes a Good Visualization?	
Textbook 2: Chapter 1, Chapter 2	
Teaching-Learning Process	<ol style="list-style-type: none"> 1. Demonstration of different data visualization tools.
Module-5	
A Deep Dive into Matplotlib	
Introduction, Overview of Plots in Matplotlib, Pyplot Basics: Creating Figures, Closing Figures, Format Strings, Plotting, Plotting Using pandas DataFrames, Displaying Figures, Saving Figures; Basic Text and Legend Functions: Labels, Titles, Text, Annotations, Legends; Basic Plots: Bar Chart, Pie Chart, Stacked Bar Chart, Stacked Area Chart, Histogram, Box Plot, Scatter Plot, Bubble Plot; Layouts: Subplots, Tight Layout, Radar Charts, GridSpec; Images: Basic Image Operations, Writing Mathematical Expressions	
Textbook 2: Chapter 3	
Teaching-Learning Process	<ol style="list-style-type: none"> 1. PPT – Comparison of plots 2. Demonstration charts
Course Outcomes	
At the end of the course the student will be able to:	
<ol style="list-style-type: none"> CO 1. Understand the data in different forms CO 2. Apply different techniques to Explore Data Analysis and the Data Science Process CO 3. Analyze feature selection algorithms & design a recommender system. CO 4. Evaluate data visualization tools and libraries and plot graphs. CO 5. Develop different charts and include mathematical expressions. 	
Assessment Details (both CIE and SEE)	
The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/course if the student secures not less than 35% (18 Marks out of 50) in the semester-end examination (SEE), and a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together	
Continuous Internal Evaluation:	
Three Unit Tests each of 20 Marks (duration 01 hour)	

1. First test at the end of 5th week of the semester
2. Second test at the end of the 10th week of the semester
3. Third test at the end of the 15th week of the semester

Two assignments each of **10 Marks**

4. First assignment at the end of 4th week of the semester
5. Second assignment at the end of 9th week of the semester

Group discussion/Seminar/quiz any one of three suitably planned to attain the COs and POs for **20 Marks (duration 01 hours)**

6. At the end of the 13th week of the semester

The sum of three tests, two assignments, and quiz/seminar/group discussion will be out of 100 marks and will be **scaled down to 50 marks**

(to have less stressed CIE, the portion of the syllabus should not be common /repeated for any of the methods of the CIE. Each method of CIE should have a different syllabus portion of the course).

CIE methods /question paper has to be designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.

Semester End Examination:

Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the subject (**duration 03 hours**)

1. The question paper will have ten questions. Each question is set for 20 marks. Marks scored shall be proportionally reduced to 50 marks
2. There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), **should have a mix of topics** under that module.

The students have to answer 5 full questions, selecting one full question from each module

Suggested Learning Resources:

Textbooks

1. Doing Data Science, Cathy O'Neil and Rachel Schutt, O'Reilly Media, Inc O'Reilly Media, Inc, 2013
2. Data Visualization workshop, Tim Grobmann and Mario Dobler, Packt Publishing, ISBN 9781800568112

Reference:

1. Mining of Massive Datasets, Anand Rajaraman and Jeffrey D. Ullman, Cambridge University Press, 2010
2. Data Science from Scratch, Joel Grus, Shroff Publisher /O'Reilly Publisher Media
3. A handbook for data driven design by Andy krik

Weblinks and Video Lectures (e-Resources):

1. <https://nptel.ac.in/courses/106/105/106105077/>
2. <https://www.oreilly.com/library/view/doing-data-science/9781449363871/toc01.html>
3. <http://book.visualisingdata.com/>
4. <https://matplotlib.org/>
5. <https://docs.python.org/3/tutorial/>
6. <https://www.tableau.com/>

Activity Based Learning (Suggested Activities in Class)/ Practical Based learning

Demonstration using projects