DATA SCIENCE

Course Code	20IT4703E	Year	IV	Semester	I
Course Category	PE-5	Branch	IT	Course Type	Theory
Credits	3	L-T-P	3-0-0	Prerequisites	Data Mining Concepts
Continuous Internal Evaluation :	30	Semester End Evaluation:	70	Total Marks:	100

	Course Outcomes	Blooms Taxonomy Level
Upon Successful completion of course, the student will be able to		
CO1	Understand the basic terms of Data Science.	L2
CO2	Understand the Data Science process.	L2
CO3	Explain how to Handle large data on a single computer	L2
CO4	Apply Data Visualization, plotting techniques.	L3

Contribution of Course Outcomes towards achievement of Program Outcomes &Strength of correlations(3:Substantial,2: Moderate,1:Slight)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3												3	
CO2	3												3	
CO3	3												3	
CO4	3												3	

Syllabus		
Unit No	Contents	Mapped CO
I	**Data science in a big data world :** Benefits and uses of data science and big data **Facets of data:** Structured data, Unstructured data, Natural language, Machine-generated data, Graph-based or network data, Audio, image, and video , Streaming data **The data science process:** Setting the research goal , Retrieving data, Data preparation, Data exploration, Data modeling or model building, Presentation and automation	CO1

II	**The data science process : Overview of the data science process**: Don't be a slave to the process, **Defining research goals and creating a project charter:** Spend time understanding the goals and context of your research, Create a project charter **Retrieving data:** Start with data stored within the company, Don't be afraid to shop around, Do data quality checks now to prevent problems later	CO1 CO2
III	**Cleansing, integrating, and transforming data:** Cleansing data, Correct errors as early as possible, Combining data from different data sources, Transforming data **Exploratory data analysis**, **Build the models:** Model and variable selection, Model execution, Model diagnostics and model comparison	CO1 CO2
IV	**Handling large data on a single computer :**The problems you face when handling large data **General techniques for handling large volumes of data:** Choosing the right algorithm, Choosing the right data structure, Selecting the right tools **General programming tips for dealing with large data sets:** Don't reinvent the wheel, Get the most out of your hardware, Reduce your computing needs.	CO1 CO3
V	**Plotting and Visualization: A Brief matplotlib API Primer**: Figures and Subplots, Colors, Markers, and Line Styles, Ticks, Labels, and Legends, Annotations and Drawing on a Subplot, Saving Plots to File, matplotlib Configuration **Plotting with pandas and sea born**: Line Plots, Bar Plots, Histograms and Density Plots, Scatter or Point Plots, Facet Grids and Categorical Data	CO1 CO4

Learning Resources
Text Books
1. Introducing Data Science: Big data, machine learning, and more, using Python tools Davy Cielen, Arno D. B. Meysman, and Mohamed Ali, Manning Publishers 2. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython Wes McKinney, Second Edition,2018,O'Reilly Media, (Unit 4- (9th Chapter)
References
1. Avrim Blum, John Hop croft and Ravindran Kannan. Foundations of Data Science. 2. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman .Mining of MassiveDatasets.v2.1,CambridgeUniversityPress.2 014.(free online) 3. Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, 3rd Edition. ISBN 0123814790.2011.
E- Resources and other Digital Material
https://www.coursera.org/browse/data-science/data-analysis https://nptel.ac.in/courses/106106179