

Year: M. Tech. I (Semester – I)

Subject Name: Data Science

Subject Code: MTCO14106

Type of course: Professional Elective - II

Prerequisite (if any): Python Programming, Statistics and Probability

List of Courses where this course will be prerequisite: --

**Rationale:** This course provides insights into the data analysis and visualization. Students can gain a thorough understanding of data science life cycle, data analytics and data visualization techniques and tools. Students will be able to solve predictive and categorization problems on real time data after learning this course.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
3	0	2	4	60	25	15	30	20	150

CA1: Continuous Assessment (assignments/projects/open book tests/closed book tests CA2: Sincerity in attending classes/class tests/ timely submissions of assignments/self-learning attitude/solving advanced problems TEE: Term End Examination TEP: Term End Practical Exam (Performance and viva on practical skills learned in course) CA3: Regular submission of Lab work/Quality of work submitted/Active participation in lab sessions/viva on practical skills learned in course

Content:

Sr.No.	Description	Total Hrs.
1	Data Science: Introduction, Life cycle, Why python, Python libraries for data science - NumPy, SciPy, Pandas, SCIKIT-Learn, IDE (Spyder) & WIDE (Web IDE- Jupyter) Data Munging: Data Types - converting, downcasting, Regex library for Strings, Aggregate, Transform, Filter, groupby, Multiindex, DateTime	5
2	Data Loading, Storage, and File Formats: What is dataset?: Reading and Writing Data in Text Format - Reading Text Files; Writing Data to Text Format; Working with Delimited Formats; JSON Data; XML and HTML: Web Scraping. Binary Data Formats - Using HDF5 Format; Reading Microsoft Excel Files; Interacting with Web APIs; Interacting with Databases	8



3	NumPy Basics: Arrays and Vectorized Computation; The NumPy ndarray; Universal Functions: Fast Element-Wise Array Functions; Linear Algebra; Pseudorandom Number Generation Pandas Basics: Dataframes Basics, Pandas Data structures -Series, Dataframe, Index Objects; creating, updating, exporting, importing, Applying functions over Series and Dataframes row wise and column wise.	7
4	Data Manipulation: Data Wrangling - Combining & Merging; Reshaping and Pivoting Data Cleaning - Missing Data-find, count, clean; Tidy Data; Exploratory Data Analysis: Descriptive Statistical Analysis- Measures for Centrality, Spread, Normality; covariance, correlation, Outlier treatment, Probability Distribution Function	8
5	Data Visualization using Graphs (Matplotlib + Seaborn): Statistical Graphics for univariate, bivariate, Multivariate, Histograms, Line Plot, Density plot, Scatter plot, Boxplot, Seaborn Themes and Styles, Network Analysis	7
6	Hypothesis testing & Data Modeling: Hypothesis testing, Two-sample testing and introduction to ANOVA, Two way ANOVA and linear regression, Modeling using regression and classification algorithms with use cases, model assessment	10

**Reference Books:**

Sr.No.	Title of book /article	Author(s)	Publisher and like details ISBN	Year of publication	Publication Edition
1.	Python for Data Analysis - Data Wrangling with Pandas, NumPy, and IPython	Wes McKinney	O'Reilly	2017	2nd Edition
2.	Pandas for Everyone - Python Data Analysis	Daniel Y. Chen	Pearson Addison Wesley -Data & Analytics Series	2020	5th Edition
3.	Mastering python for data science	Samir Madhavan	Ingram short title, ISBN : 978-1784390150	2015	1st Edition

4.	Applied Data Science with Python and Jupyter	Alex Galea	Packt Publications, ISBN : 978-1-78995-817 -1	2018	1st Edition
5.	Data Science for Dummies	Lillian Pierson	by John Wiley & Sons, ISBN : 978-1-119-32763-9	2017	2nd Edition

**Course Outcomes:**

Sr. No.	CO statement	Marks % weightage
CO-1	Demonstrate how to generate, collect, clean and process data.	10%
CO-2	Conduct exploratory data analysis using statistics and relevant visual methods.	20%
CO-3	Solve and implement real-world data science applications using various common Python libraries.	15%
CO-4	Distinguish among common visual presentations of data used for effective communication.	15%
CO-5	Describe applicability of data science to solve real world problems	20%
CO-6	Utilize different machine learning algorithms to model data.	20%

**List of Open learning website:**

1. NPTEL Course on Data Analytics using Python :  
[https://onlinecourses.nptel.ac.in/noc21\\_cs45/course](https://onlinecourses.nptel.ac.in/noc21_cs45/course)
2. NPTEL Course on Python for Data Science : [https://onlinecourses.nptel.ac.in/noc21\\_cs33/](https://onlinecourses.nptel.ac.in/noc21_cs33/)
3. Coursera Specialization Course on Applied Data Science with Python Programming :  
<https://www.coursera.org/specializations/data-science-python>
4. Course on Data Science with Python : <https://www.simplilearn.com/getting-started-data-science-with-python-skillup>

**List of Open Source Software:**

1. Open Source Python Libraries : NumPy, SciPy, Pandas, SCIKIT-Learn, TensorFlow, BeautifulSoup, Matplotlib, Seaborn



2. IDEs: Jupyter Notebook, Spyder, WIDE -Google Colab

**FOR LAB SESSIONS:**

**List of Experiments:**

Sr. No.	Practical
1	Learn features and functions to manage and analyze large amounts of data with Excel.
2	Write a program to collect the Twitter data using Twitter API and Python libraries.
3	Install Anaconda, practice basic commands and various IDEs.
4	Manipulate arrays in Python using Numpy.
5	Practice data manipulation with Pandas. (Practice data series, data frames, data selection, sorting, searching and statistics with Pandas Python library.)
6	Practice different machine learning algorithms and statistical modeling using Scikit-Learn.
7	Build machine learning model using Scikit-Learn to perform sentiment analysis. Use any recent text dataset from UCI, GitHub, Kaggle or similar platform.
8	Create interactive data visualizations using Matplotlib and Seaborn.
9	Practice Single and Multivariate regression model with dataset.
10	Practice classification modeling using machine learning with TensorFlow and Keras. (Hint : Explore TensorFlow Python library to build Machine Learning models.)

