Year: B. Tech III (Semester VI)

Subject Name: Python for Data Science
Type of course: Professional Core Course
Prerequisite (if any): Python Programming

Science Subject Code: BTCO13602

List of Courses where this course will be prerequisite: Machine Learning, Deep Learning, Data Analytics

Rationale: Learners can gain a thorough understanding of data analytics methods and techniques through this Data Science with Python program. Data analysis, visualization, NumPy, SciPy, web scraping, and text processing are all skills you can learn with Python. This course provides insights into the data analysis and visualization. Students will be able to perform predictive and categorization modeling on real time data.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
4	0	2	5	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 Hours
Unit 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	8
Unit 2	Data Understanding : 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	10
Unit 3	NumPy Basics: Arrays and Vectorized Computation: The NumPy ndarray; Universal Functions: Fast Element-Wise Array Functions; Matrix and	10

	associated operations, Linear Algebra; Pseudorandom Number Generation Pandas Basics: Dataframes Basics on some dataset, Pandas Data structures - Series, Dataframe, Index Objects; creating, updating, exporting, importing, Applying functions over Series and Dataframes row wise and column wise.	
Unit 4	Exploratory Data Analysis: Descriptive Statistical Analysis- Measures for Centrality, Spread, Normality; covariance, correlation, Outlier treatment, Probability Distribution Function	10
Unit 5	Data Visualization using Graphs Statistical Graphics for univariate, bivariate, Multivariate, Histograms, Line Plot, Density plot, Scatter plot, Boxplot, Pair plot, Seaborn Themes and Styles, Network Analysis	10
Unit 6	Data Modeling: Linear Model(Simple Linear Regression, Multiple Linear Regression, Logistic Regression Model using sklearn, statsmodel), Classification algorithm CASE STUDY: Regression (Predicting price of pre-owned cars or other) Classification (Classifying personal income or other)	12

Suggested Specification table with Marks (Theory): (For B.Tech only)

Distribution of Theory Marks					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
3	20	30	7	0	0

Legends: R: Remembrance, U: Understanding; A: Application, N: Analyze, E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Reference Books:

Sr No	Title of book /article	Author(s)	Publisher and details like ISBN	Year of publication / Publication Edition
1.	Python for Data Analysis - Data Wrangling with Pandas, NumPy, and IPython	Wes McKinney	O'Reilly	Latest Edition

2.	Pandas for Everyone - Python Data Analysis	Daniel Y. Chen	Pearson Addison Wesley -Data & Analytics Series	
3.	Mastering python for data science	Samir Madhavan	Ingram short title, ISBN : 978-1784390150	
4.	Applied Data Science with Python and Jupyter	Alex Galea	Packt Publications, ISBN : 9781-789958171	
5.	Data Science for Dummies	Lillian Pierson	by John Wiley & Sons, ISBN : 978-1-119-32763-9	

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
2. NPTEL COurse on Python for Data Science : 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. Skill Up by SimplyLearn 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.	Practicals
1	Practice Python basic concepts such as data types and variables, control structures, functions
2	Install Anaconda, practice basic commands and various IDEs.
3	Manipulate arrays in Python using Numpy.
4	Practice data manipulation with Pandas. (Practice data series, data frames, data selection, sorting, searching and statistics with Pandas Python library.)
5	Practice url parsing with custom HTML and XML parsing.
6	Practice real time HTML and XML web scraping.
7	Practice machine learning and statistical modeling using Scikit-Learn.
8	Create interactive data visualizations using Matplotlib & 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.)