

Year: B. Tech III (Semester VI)

Subject Name: Python and SQL for Data Science

Subject Code: BTEA19621

Type of course: Honors (Group: Data Science)

Prerequisite (if any): Machine learning and Deep learning

Rationale: Students will be able to learn and practise using different Python libraries for creating real-time data science projects in this course. After completing this course, students will be able to create and query tables in relational databases such as MySQL. Students in this course will have several hands-on labs and assignments to help them gain practical knowledge and expertise with all required libraries and data sets.

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.	Contents	Total Hrs
1.	Python Libraries for Data Science : NumPy, SciPy, Pandas, SciKit-Learn, Machine Learning Libraries - Tensorflow, Keras	08
2.	NumPy Basics: Arrays and vectorized computation; The numpy ndarray; universal functions: Fast Element-wise array functions; matrix and associated operations, linear algebra; pseudorandom number generation Scientific Computing using SciPy : Working with sparse, spatial data, optimizer, interpolation	10
3.	Pandas Basics: Pandas data structures -series and dataframes Data pre-processing using Pandas: Indexing objects, creating, updating, exporting, importing, applying functions over series and data frames - row wise and column wise, slicing, filtering. Case Study : Analyze appropriate dataset using pandas	10
4.	Data Modeling using Tensorflow, Keras and SciKit - Learn: Tensorflow dataflow graphs, tensorflow elements - constant, variable, place holders, sessions, program execution phases, logging and visualizing activities using tensorboard, elements of Keras, model creation using Keras - sequential and functional, model compile and execution	14



	Simple neural network, linear model (Simple Linear Regression, Multiple Linear Regression, Logistic Regression Model using sklearn, statsmodel), classification algorithm	
5.	Accessing Databases using Python: Basic concepts related to using Python to connect to databases, create tables, load data, query data using SQL, and analyze data using Python	09
6.	Data Visualization using Seaborn Visualization for univariate, bivariate, multivariate, histograms, line Plot, density plot, scatter plot, boxplot, pair plot, seaborn themes and styles	09

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
5	5	5	15	10	20

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

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.	Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems	Aurelien Geron	Ingram short title, ISBN : 9789352139057, 9352139054	
4.	Python Data Science Essentials	Alberto Boschetti, Luca Massaron	9781789537864	
5.	Core Python Programming	Nageshwara Rao	DreamTech	



6.	MySQL for Python: Database Access Made Easy	<u>Albert Lukaszewski</u> PhD	Packt 1849510180
----	---	----------------------------------	---------------------

Note: Students should refer to the latest editions of books

Course Outcomes:

Sr. No.	CO statements	Marks % weightage
CO-1	Select appropriate python library to render the multidimensional data.	20%
CO-2	Create Pandas dataframes by generating, collecting, cleaning and processing data.	20%
CO-3	Produce various graphs using Seaborn library	20%
CO-4	Solve and implement real-world data modeling problems using tensorflow, keras and scikit-learn.	25%
CO-5	Create and Integrate the database with python for data science project	15%

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. <https://www.udemy.com/course/sql-for-data-science/>

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	Render multidimensional data using Numpy.

4	Implement various data preprocessing methods using pandas.
5	Create and simulate following expressions considering tensors of various dimensions and different data types such constant, variable and place holder : 1. $22x^2 + 2x + 3 = 0$ 2. $14x^3 + 7x^2 + 3x + 4 = 0$ Also visualize the computational graph using a tensorboard.
6	Create a simple neural network using sequential and functional models of keras. Visualize the computational graph using tensorboard.
7	Practice K-Means and K-Medoid Clustering algorithm of machine learning using Scikit-Learn.
8	Implement Decision Tree Classifier of machine learning using Scikit-Learn.
9	Create various graphs using advanced features of seaborn for an appropriate dataset.
10	Mini Project to apply various python libraries and access MySql database for data science project.

