SARVAJANIK UNIVERSITY
INCLUSIVE | INTEGRATED | INNOVATIVE

SARVAJANIK UNIVERSITY
Sarvajanik College of Engineering and Technology
Department of Computer Engineering / Information Technology

Ph.D Semester 1
Course Work

Subject Name: Python Skills for Research **Subject Code:** PTCO13101
Type of course: Course Work for PhD
Prerequisite (if any): Basics of Python Programming, AI techniques

Rationale: This course provides insight into advanced concepts that are useful to researchers when developing python code that is robust, highly optimized, efficient and normalized. Students will be able to learn advanced Python programming and various modules and packages of python which are useful for research in detail.

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.	Content	Total Hrs
1	**Python Modules and Packages :** Built-in Modules - Importing Modules in Python Programs, Working with Random Modules, e.g. - builtins, os, time, datetime, calendar, sys, etc., User Defined Functions - Structure of Python Modules., object oriented Programming, high order functions - map, reduce, filter	3
2	**Working with files:** CSV, JSON, Excel, Text files	5

	Working with Database: Introduction to MySQL, PYMYSQL Connections, Executing queries, Transactions, Handling error.	
3	**Data Visualization: Seaborn**-Style functions, Color palettes, Distribution plots, Categorical plots, Regression plots, Axis grid objects, **Tableau** - Modify data sources, build advanced charts, use complex calculations	6
4	**Python Research Tools:** Introduction to Python modules commonly used in scientific computation **NUMPY:** Creating NumPy arrays, Indexing and slicing in NumPy, Downloading and parsing data, Creating multidimensional arrays, NumPy Data types, Array attributes Indexing and Slicing, Creating array views copies, Manipulating array shapes I/O **SCIPY:** Introduction to SciPy, Create function, modules of SciPy **PANDAS:** Using multilevel series, Series and Data Frames, Grouping, aggregating, Merge DataFrames, Generate summary tables, Group data into logical pieces, Manipulate dates, Creating metrics for analysis, Data wrangling, Merging and joining, Analytics Vidhya dataset- Loan, Prediction Problem, Data Mugging using Pandas, Building a Predictive Model **WEB SCRAPING:** Scraping Webpages, Beautiful soup package, Real time project **SCIKIT-LEARN, Keras :** Classification, Clustering, Dimensionality reduction	12
5	Statistical Learning: Exploration of statistical learning using the scikit-learn library with case study. 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 - **Supervised learning :** Regression - Different types of Regression, Linear Regression, Logistic Regression,	11

	Decision tree Algorithms, Classification - Naive- Bayes', KNN Classification, Curse of Dimensionality-PCA, SVM Classification **Unsupervised learning :** Clustering - k-means clustering, Random Forest	
6	Time Series Analysis and Forecasting in Python- importing, visualizing time series data **Image processing :** Introduction to digital image processing, Basic operations on an image – Crop, Scale, Rotate, Flip, Changing contrast, brightness and color, Edge detection, blur, sharpening. NLP: Tokenisation, stemming, lemmatization, stop word removal, unigram, bigram, ngram, sentence segmentation Case Studies	8

Reference Books:

Sr No	Title of book /article	Author(s)	Publisher and details like ISBN	Year of publication	Publication Edition
1	Introduction to Computation and Programming Using Python	John V Guttag	Prentice Hall of India	2016	2nd Edition
2	Core Python Programming	R. Nageswara Rao	dreamtech	2021	3rd Edition
3	Pandas for Everyone - Python Data Analysis	Daniel Y. Chen	Pearson Addison Wesley -Data & Analytics Series	2020	5th Edition
4	Programming Python	Mark Lutz	O`Reilly	2010	4th Edition
5	Python for Data Analysis	Wes McKinney	O'Reilly	2017	2nd Edition

Course Outcomes:

Sr. No.	CO statement	Marks % weightage
CO-1	Utilize python for connection to database for applications using various modules.	20
CO-2	Distinguish among common visual presentations of data used for effective communication.	15
CO-3	Apply different machine learning algorithms to model data.	25
CO-4	Demonstrate time-series analysis, image and text processing in various applications	25
CO-5	Solve and implement real-world data science applications using various common Python libraries.	20

List of Open learning website:

1. NPTEL Course on Data Analytics using Python : ttps://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. 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:

SARVAJANIK UNIVERSITY
Sarvajanik College of Engineering and Technology
Department of Computer Engineering / Information Technology

List of Experiments:

Sr. No	Practical
1	Install Anaconda, practice basic commands and various IDEs.
2	Develop programs to understand the control structures of python
3	Develop programs to learn different types of structures (list, dictionary, tuples) in python
4	Develop programs to learn concept of functions scoping, recursion and list mutability.
5	Develop programs to learn regular expressions using python.
6	Learn features and functions to manage and analyze large amounts of data with Excel.
7	Manipulate arrays in Python using Numpy.
8	Practice data manipulation with Pandas. (Practice data series, data frames, data selection, sorting, searching and statistics with Pandas Python library.)
9	Create interactive data visualizations using Matplotlib and Seaborn.
10	Implement Logistic Regression in Python using machine learning for the Social_Networks_Ads.csv dataset.
11	Implement the following Ensemble algorithms on the Iris dataset or the labelled dataset of your choice. Prepare a comparison table of accuracy measures for different number of estimators for the given algorithms. 1. Random Forest 2. AdaBoost 3. Gradient Boost
12	Calculate the Cost metrics and/or the Cost function for the following datasets after preprocessing if necessary. Further, use cross validation and record the results. 1. Linear Regression on Boston Housing dataset available at : https://www.kaggle.com/schirmerchad/bostonhoustingmlnd#housing.csv 2. Classification or Logistic Regression on following datasets: Coronavirus patients.csv at https://www.kaggle.com/kimjihoo/coronavirusdataset
13	Create interactive data visualizations using Matplotlib and Seaborn.
14	Practice classification modeling using machine learning with TensorFlow and Keras.