SDSC2001: PYTHON FOR DATA SCIENCE

Effective Term
Semester A 2023/24

Part I Course Overview

Course Title
Python for Data Science

Subject Code
SDSC - School of Data Science
Course Number
2001

Academic Unit
School of Data Science (DS)

College/School
School of Data Science (DS)

Course Duration
One Semester

Credit Units
4

Level
B1, B2, B3, B4 - Bachelor's Degree

Medium of Instruction
English

Medium of Assessment
English

Prerequisites
CS2311/CS2315 Computer programming / CS1315 Introduction to Computer Programming and SDSC1001 Introduction to Data Science*

* Pre-requisite SDSC1001 will be exempted for students who are enrolled in Minor in Data Science

Precursors
Nil

Equivalent Courses
Nil

Exclusive Courses
Nil

Part II Course Details

Abstract

This course provides students with an extensive exposure to the use of Python specifically for data science. Topics include Python language fundamentals, data analysis using Python libraries, applied machine learning and text mining in Python, and the practice of scientific computing. The students will acquire hands-on experience using Python and the popular packages related to data manipulation, processing and analysis, with the minimal backgrounds of theories in methodology aspects. The students are required to be able to independently develop Python codes from scratch in professional way for elementary algorithms in data sciences.

Course Intended Learning Outcomes (CILOs)

	CILOs	Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Familiarize with basic Python programming language, the version-control tool like github, and the pre-processing of large datasets.	20	x		
2	Use Python libraries for basic data representation, manipulation, mining, analysis and visualization.	20	x	x	
3	Implement algorithms for fundamental methods in data analysis	30	x	x	x
4	Use Python for exploratory data analysis to solve some practical problems	30	x	x	x

A1: Attitude
Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability
Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to real-life problems.

A3: Accomplishments
Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

Teaching and Learning Activities (TLAs)

	TLAs	Brief Description	CILO No.	Hours/week (if applicable)
1	Lecture	Learning through teaching is primarily based on lectures.	1, 2, 3, 4	3 hours per week
2	Tutorial	Using the computers in the lab to practice the methods taught in lecture	2, 3, 4	1 hour per week

| 3 | Take-home assignments | Learning through take-home assignments is primarily based on interactive problem solving and hand-on computer exercises allowing instant feedback. | 2, 3, 4 | after class |

Assessment Tasks / Activities (ATs)

	ATs	CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)
1	Hand-in assignments	2, 3, 4	25	The assignments enable students to demonstrate their Python programming skills and understanding of concepts and methods for data science.
2	Course Project	2, 3, 4	25	The course project provides students chances to demonstrate their achievements on practical use of data analysis methods learned in this course for advanced data science topics.

Continuous Assessment (%)

50

Examination (%)

50

Examination Duration (Hours)

2

Additional Information for ATs

Note: To pass the course, apart from obtaining a minimum of 40% in the overall mark, a student must also obtain a minimum mark of 30% in both continuous assessment and examination components.

Assessment Rubrics (AR)

Assessment Task

Hand-in assignments

Criterion

1.1 Ability to learn the basic concepts and skills of Python language, and to write computer programs with the aid of common Python libraries for data representation, manipulation, analysis and visualization.

Excellent (A+, A, A-)

High

Good (B+, B, B-)
Significant

Fair (C+, C, C-)
Moderate

Marginal (D)
Basic

Failure (F)
Not even reaching marginal levels

Assessment Task
Hand-in assignments

Criterion
1.2 Capability to apply popular machine learning Python libraries for data analysis.

Excellent (A+, A, A-)
High

Good (B+, B, B-)
Significant

Fair (C+, C, C-)
Moderate

Marginal (D)
Basic

Failure (F)
Not even reaching marginal levels

Assessment Task
Course Project

Criterion
Ability to develop, validate and publish Python code to solve real problems by using elementary data analysis methods.

Excellent (A+, A, A-)
High

Good (B+, B, B-)
Significant

Fair (C+, C, C-)
Moderate

Marginal (D)
Basic

Failure (F)
Not even reaching marginal levels

Part III Other Information

Keyword Syllabus

Basics of Python programming language:
Installation and setup of Python and Python Packages (shell and interactive environment); IPython and Jupyter Notebook; basics of Python language; data structures in Python; file operations; functions in Python; OOP; data processing; profiling and timing; version control skills (github)

Introduction to Python for scientific computing, data processing and plotting:
Scientific computing with NumPy (multidimensional array, indexing and slicing, array functions, pseudo-random generators); data manipulation with Pandas (series, dataframe, selection, filtering, sorting, rankding); visualization with Matplotlib.

Python for elementary machine learning (Scikit-Learn):
Pre-processing; feature extraction; simple liner regression analysis; binary classification analysis; decision tree and random forest; k nearest neighbour; model selection.

Reading List

Compulsory Readings

	Title
1	Python Data Science Handbook: Essential Tools for Working with Data, by Jake VanderPlas, O'REILLY 2016

Additional Readings

	Title
1	Data Analysis from Scratch with Python: Beginner Guide for Data Science, Data Visualization, Regression, Decision Tree, Random Forest, Reinforcement Learning, Neural Network and NLP using PythonBy Peter MorganPublisher: AI Sciences LLC; 2 edition (August 13, 2018)
2	Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython 2nd EditionBy Wes McKinney O'Reilly Media; 2 edition (September 25, 2017)