Fall 2023 CS 677 A2
Data Science with Python
Tuesdays 6:00-8:45 pm, CAS 426

Joshua Enxing
jenxing@bu.edu
Office Location: N/A
Office Hours: Saturdays 2-4 pm on Zoom (link on blackboard)

Course Structure: Lectures will be given in-person during each course period. We will usually have a ten- or fifteen-minute break halfway through the class. Please bring your computers, as there will be in-class Python exercises we will do (if you do not have access to a portable personal computer, please reach out to me).

Course Description: At the present time, there is a growing need for specialists with background in Python who can apply data science methods to practical problems at their workplace. Working in data science requires an understanding of many interdisciplinary concepts, involves data mining and application of various methods.

The course is designed to fill this need. Students will learn major Python tools and techniques for data science and machine learning. There are assignments on topics covered in class every two weeks. These assignments will help students build necessary statistical, visualization and other data science skills for effective use of data science in a variety of applications including finance, text processing, time series analysis and recommendation systems. In addition, students will choose a topic in data science for a final project.

Prerequisite(s): MET CS 521 or equivalent (or instructor's consent). The course can be taken by students with not exclusively computer science backgrounds who have basic knowledge of Python.

Required text: *Python Data Science Handbook*; Vanderplas, Jake; O'Reilly Publishing; 2017 Edition (ISBN 13: 9781491912133)

Other materials:

- *Class Slides and any supplemental materials posted on blackboard*, required.
- **Recommended text:** *Mathematics for Machine Learning*; Deisenroth, Marc Peter / Faisal, A. Aldo / Ong, Cheng Soon; Oxford Publishing; 2020 Edition (ISBN 13: 9781108455145)

Computer Software: We will use Python for all course examples and assignments. In class, I will usually use Google Colaboratory notebooks since the modular nature makes them better for pedagogical applications. You are welcome to use any IDE you like, or Google Colab as well.

Course Materials & Website: Announcements, assignments, course notes, and other course information will be posted on Blackboard.

Class Environment: I want us to have a safe, enjoyable learning environment that reflects a diversity of perspectives, experiences, and identities. If you feel any aspect of the class is not supportive, please let me know.

Learning Objectives: By successfully completing this course, you will be able to:

- Use different features of Python and Numeric Python
- Design and develop applications using pandas and graphics, data scaling, error analysis
- Design and develop applications with logistic regression and nearest neighbor classification
- Design and develop applications with linear models and regression.
- Design and develop applications using decision trees, random forests, and naive bayes
- Design and develop applications using support vector machines and k-means clustering

Assessment: Work for the course (and the components of your final grade) will include:

- **Quizzes**: There will be one half-hour long in-person quiz approximately every two weeks (see tentative class schedule). These will be closed-book and closed-notes and will not require the use of a computer.
- **Assignments in Python**: There will be six programming assignments with specific due dates that you will be required to submit on Blackboard (see tentative class schedule). Any late assignments will be docked 20% every day (rounded up to the nearest whole number) that they are late. Assignments will be made available two weeks before the due dates.
- **Final Exam**: There will be an in-person final exam held on 12/19 at 6:00 pm EST in CAS 426. As with the quizzes, this exam will be closed-book and closed-notes and tests you on theoretical material not requiring the use of a computer (more details will be provided on Blackboard closer to the time of the exam).
- **Final Project**: This course will culminate in a final presentation to polish your speaking skills. The project will be individual and will consist of a coding portion as well as a recorded five-minute presentation portion. You will be required to submit a project proposal approximately a month before the project due date (see tentative schedule) to make sure your proposed topic/method is appropriate (i.e. not too difficult or will not involve more work than you plan). More detail will be given about the project on Blackboard.

Grading:

Quizzes	20%
Assignments in Python	35%
Final Exam	30%
Final Project	15%
Total	100%

Accommodations for Students with Disabilities: In accordance with University policy, every effort will be made to accommodate students with respect to speech, hearing, vision, or other disabilities. Any student who may need an accommodation for a documented disability should contact Disability and Access Services (https://www.bu.edu/disability/) at 617-353-3658 or at mailto:access@bu.edu for review and approval of accommodation requests. Once a student receives their accommodation letter, they must send it to their instructor each semester. **Accommodations cannot be implemented if the student does not send their letter.**

Academic Integrity: Please visit
https://www.bu.edu/met/current-students/academic-policies-procedures/ for the full Metropolitan College Academic Conduct Code. By registering for this course, you agree to abide by the policies outlined here. Any violations of this code will be reported to the Dean's office.

Attendance Policy: While there is no attendance policy for class, if you miss a class where there is a quiz without there being extenuating circumstances, you will not get credit for that quiz.

Tentative Course Schedule
All assignments are due on Blackboard at 11:59 pm EST on the day listed. Chapter number refers to the chapter of VanderPlas (in addition, material will be posted on Blackboard for that topic, in the form of Modules and presentation slides and possible links to other sites or excerpts from other textbooks).

Week (Date)	**Topics**	**Chapter**	**Assignments Due**
1 (9/5)	Review of Python, Numpy	2	
2 (9/12)	Numpy (cont.), Pandas	2, 3	Quiz 1
3 (9/19)	Pandas (cont.), Matplotlib	3, 4	Assignment 1
4 (9/26)	Intro to Machine Learning	5	Quiz 2
5 (10/3)	K-NN, Logistic Regression		Assignment 2
6 (10/10)	**No class, BU Monday**		
7 (10/17)	Probability, Distributions, other math		Quiz 3
8 (10/24)	Linear Regression	5	Assignment 3
9 (10/31)	Naive Bayes	5	Quiz 4
10 (11/7)	Decision Trees/Random Forests	5	Assignment 4
11 (11/14)	Support Vector Machines	5	Quiz 5, **Project Proposal**
12 (11/21)	Unsupervised Learning	5	Assignment 5
13 (11/28)	Feature Engineering	5	Quiz 6
14 (12/5)	Principal Component Analysis	5	Assignment 6
15 (12/12)	Final Exam Review		**Final Project**
16 (12/19)	**Final Exam, 6:00-8:30 pm ET, CAS 426**		

Disclaimer: The Instructor reserves the right to make changes to the syllabus as necessitated by circumstances. At times, depending on the teaching pace, I may be ahead or behind. Any changes in dates will be announced on Blackboard and in class.