MSA 8010: Data Programming (Section 008)

Syllabus for Fall 2023*

Instructor: Xiangshi Yin†

August 27, 2023

Contents

*The course syllabus provides a general plan for the course; deviations may be necessary.

†Email: xyin@gsu.edu

1 Description

This course introduces common Python programming practices for data analysis. The objective is to prepare students with programming skills to tackle large data analysis projects.

1.1 Instructor

Name	Xiangshi Yin
Email	xyin@gsu.edu

1.2 Teaching Assistant

Name	Dixsha Mahna
Email	dmahana1@gsu.edu

1.3 Lectures

We meet on every Tuesday evening at 7:15 PM (Eastern Time).

Days	Tuesday
Time	7:15 PM - 9:45 PM (Eastern Time)
Room	GSU Buckhead Executive Ed Ctr **Room: 1216**

To access the class materials online, you need to:

- Go to the class home page on iCollege at `https://gastate.view.usg.edu/d2l/home/2886149`
- Homework assignments and quizzes will be posted
- Please follow the announcements page for any course udpates

1.4 Office Hours

Days	Tuesday
Type	By Appointment
Time	TBD[1]
Room	TBD[2]

[1] Will be available in week 2
[2] Most likely in the hoteling office of the Buckhead building, will confirm before week 2

There is a 30 minutes session available every Tuesday before our regular classes. To book the time on calendar, you need to:

- Use this link `https://outlook.office.com/bookwithme/user/ba5e4b49e7d9445daf7923151dd343c6@gsu.edu?anonymous&ep=plink`
- Choose the preferred time slot

1.5 Contact the instructor

During the term, it is highly recommended that you contact the instructor, via scheduled office hours or via email. They are available to help you focus your projects, gain access to resources, and answer your questions. Please try to contact them at least once during the term to discuss your project. Your class members are also a good source of help.

1.6 Course Web-site

Class information will be posted on the iCollege site. There will be links to other web-sites with course related material.

2 Overview

The goal of this course is to prepare students with Python programming skills to handle data analysis projects. In the first half of the course, we'll cover topics on basic Python coding skills, and tools for common use cases of data analytics and business reporting. In the second half of the course, we'll be talking about some advance topics on machine learning modeling from the application perspective with Python programming.

2.1 Intended audience

The course is aimed to students who are beginners in programming and Python language. Basic knowledge on Statistics and Linear Algebra would be helpful to understand the machine learning topics we'll be talking about in the second half of the course.

2.2 Learning objectives

Upon successful completion of this course, you will accomplish the following objectives and outcomes. In particular, students who complete this course will gain "Ready for work" skills, including:

- Explore, analyze and manipulate data sets
- Prepare features sets for modeling
- Apply machine learning tools and techniques
- Interpret and present results

3 Schedule

The course schedule[3] is shown below. However, the topics and readings may change according to the interests and abilities of the class. Please refer to course content page on iCollege for most updated lecture guidelines. Materials may be updated 24 hours prior to class; please check before attending class.

Session	Date	Topic	Reading
01	2023-08-22	Python Crash Course (p1)	PDA[4] Chapter 1-3
02	2023-08-29	Python Crash Course (p2)	PDA Chapter 3
03	2023-09-05	Python Crash Course (p3)	PDA Chapter 3
04	2023-09-12	Numpy and Linear Algebra	PDA Chapter 4
05	2023-09-19	Pandas and Dataframe	PDA Chapter 5
06	2023-09-26	Pandas Data Table Exercise	PDA Chapter 6-8
07	2023-10-03	Statistics with Python	PDA Chapter 6-8
08	2023-10-10	Data Visualization	PDA Chapter 9
09	2023-10-17	Data Visualization and Regression Analysis	HML[5] Chapter 1, FML[6] Chapter 2&8
10	2023-10-24	Introduction to Machine Learning (p1)	HML Chapter 2, FML Chapter 3
11	2023-10-31	Introduction to Machine Learning (p2)	HML Chapter 6-7, FML Chapter 4
12	2023-11-07	Information based learning	FML Chapter 4
13	2023-11-14	Unsupervised Learning Model	HML Chapter 9, FML Chapter 6
14	2023-09-19	*Thanksgiving Break* (**No Classes**)	N/A
15	2023-11-21	Project Presentation	N/A

4 Readings

4.1 Books

Primary Textbook:

- **Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter** 3rd Edition by Wes McKinney
  - Released August 2022
  - Publisher(s): O'Reilly Media, Inc.
  - ISBN: 9781098104030
- **Hands-On Machine Learning with Scikit-Learn and TensorFlow** O'Reilly Media, 3rd Edition by Aurélien Géron
  - Released October 2022
  - Publisher(s): O'Reilly Media, Inc.
  - ISBN: 9781098125974

[3]The course syllabus provides a general plan for the course; deviations may be necessary.
[4]Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter
[5]Hands-On Machine Learning with Scikit-Learn and TensorFlow
[6]Fundamentals of Machine Learning for Predictive Data Analysis

- **Fundamentals of Machine Learning for Predictive Data Analytics** by John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy
  - Hardcover: ISBN 9780262029445, 624 pp., July 2015
  - eBook: ISBN 9780262331722, 624 pp., July 2015

4.2 Other books and resources

- Fabrizio Romano *Learning Python: Learn to code like a professional with Python - an open source, versatile, and powerful programming language* Packt Publishing, 2015.
- Charles Severance *Python for Everybody: Exploring Data in Python 3* CreateSpace Independent Publishing Platform, 2016
- Joel Grus *Data Science from Scratch: First Principles with Python* O'Reilly Media, 2015.
- Foster Provost *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking* O'Reilly Media, 2013.
- *An Introduction to Statistical Learning, with Application in Python* `https://hastie.su.domains/`
- Technical blogs posted on `www.medium.com`.
- Last but not least, Google and StackOverflow are always your BEST FRIENDS to learn special coding skills.

5 Software

- All programming activities will be performed on the your own laptop. Your laptop should have Python 3 and Jupyter Notebook installed. Using the Anaconda installation (`https://www.anaconda.com/products/individual`) is a good start to have most of the packages we need for the class in one shot. Detailed installation instructions will be posted on the class home page on iCollege.
- If you are interested to explore new tools, you could also try Google Colab (`colab.research.google.com`). It is an online Jupyter Notebook environment with Python and free computing resources backed by Google. You may need to install certain packages yourselves if they are not available in the notebook environment.

6 Homework/Quizzes/Final Project

There will be 5 quizzes, 5 home assignments, and 1 final group project over the whole semster. Detailed due dates are listed below.

Date	Note	Quiz	Homework	Quizz/Homework Due(EST)
2023-08-22	First Class			N/A
2023-08-29		Y		2023-09-05 19:15
2023-09-05			Y	2023-09-12 19:15
2023-09-12		Y		2023-09-19 19:15
2023-09-19			Y	2023-09-26 19:15
2023-09-26	Team assigned	Y		2023-10-03 19:15
2023-10-03	Team finalized		Y	2023-10-10 19:15
2023-10-10	Suggested project topics	Y		2023-10-17 19:15
2023-10-17			Y	2023-10-24 19:15
2023-10-24	Project topics finalized	Y		2023-10-31 19:15
2023-10-31			Y	2023-11-07 19:15
2023-11-07				N/A
2023-11-14				N/A
2023-11-21	Thanksgiving Break (**No Classes**)			N/A
2023-11-28	Project Presentation			2023-12-08 23:59

6.1 Homework

Homework assignments are the continuation of a hands-on activities in class. Detailed information about the activity and expectation for successful completion are provided with the instructions. See the web site for the most recent and detailed information on these assignments. **Homeworks are individual assignments!** You may discuss the assignment with your classmates, but your final answers should reflect your individual effort. Completed assignments must be uploaded[7] by the deadline.

6.2 Quizzes

Quizzes will be given out every 2-3 weeks after the class and comprise only a few questions. However, some questions may need some thinking and calculations.

6.3 Final Group Project

The project has to showcase a subset of the methods and tools that are introduced in this course. Teams can comprise up to 2 students, and should form within the first few weeks of the term. **Teams are free to choose a data set for their project** (the instructor will also release a list of suggested topics as a reference). The use of proprietary or classified data sets is not allowed. **Project deliverables include a detailed report, functioning code, and a presentations.** Details about requirements and evaluation criteria will be posted on the class homepage on iCollege.

Key dates to remember (also marked in the schedule of the previous page):

- 2023-09-26: Project teams will be assigned[8]
- 2023-10-03: Project team will be finalized[9]

[7] Instructions on homework submission will be posted on the class home page on iCollege
[8] The instructor will assign students into project teams
[9] Exception: you may withdraw from a team at any time afterwards and submit the project assignment individually.

- 2023-10-10: The instructor will release a list of suggested project topics
- 2023-10-24: Team leads need to submit their teams' choice on final project topics
- 2023-11-28: Final project presentation
- 2023-12-08: Report and code due

Teams will submit one assignment for all team members. In most cases, each member of the team will get the same score. Each team assignment must also include a list of tasks completed by each member.

7 Evaluation

Students will be evaluted by the deliverables summarized below:

Assignment	Percentage
Quizzes	20%
Homework	40%
Final Project	40%
Total	100%

Grade	Percentage
A+	≥ 97
A	≥ 90
A-	≥ 87
B+	≥ 83
B	≥ 80
B-	≥ 77
C+	≥ 73
C	≥ 70
C-	≥ 67
D	≥ 60
F	< 60

8 Workload Expectations

Students should plan for 2 - 3 hours of work outside of class each week for each course credit hour. Thus, a 3-credit course averages between 6 and 9 hours of student work outside of the classroom, each week.

Arbitration: There will be a one-week arbitration period after graded activities are returned. Within that one-week period, you are encouraged to discuss any assumptions and/or misinterpretations that you made on the activity that may have influenced your grade.

Attendance: If you are unable to attend a class session, it is your responsibility to acquire the class notes, assignments, announcements, etc. from a classmate. The instructor will not give private lectures for those that miss class.

Submission of Deliverables: Unless specific, prior approval is obtained, no deliverable will be accepted after the specified due date. If you have a legitimate personal emergency (e.g., health problem) that may impair your ability to submit a deliverable on time, you must take the initiative to contact the instructor before the due date/time (or as soon after your emergency as possible) to communicate the situation.

9 General GSU Policies

9.1 Students with disabilities or special needs

Students who wish to request an accommodation for a disability may do so by registering with the GSU Access and Accommodations Center (AACE). Students may only be accommodated upon issuance by the AACE of a signed Accommodation Plan and are responsible for providing a copy of that plan to instructors of all classes in which accommodations are sought. Please let me know if you have a disability or special need that requires accommodation.

9.2 Religious accommodation and holidays

Students must provide instructors with reasonable notice of the dates of religious holidays on which they plan to be absent and must be given an equivalent opportunity to make up missed work according to an agreed-upon schedule. Such accommodations might include rescheduling an exam or giving the student a make-up exam, allowing an individual or group presentation to be made on a different date, letting a student attend a different section for the same class that week, adjusting a due date or assigning the student appropriate make-up work that is no more difficult than the original assignment. Students wishing to have an excused absence due to the observation of a religious holiday of special importance must provide **"advance written request to each instructor by the end of the first week of classes."**

9.3 GSU policy on withdrawing from classes

The semester midpoint (October 10, 2023) is the last day to voluntarily withdraw from a full semester class and receive a possible grade of W. Withdrawals appear on the student's permanent record and count toward their attempted hours.
Students can use PAWS to withdraw before the midpoint; after that date, voluntary withdrawals cannot occur. Students are allowed only 6 withdrawals during their academic careers at GSU. If they withdraw from your course after drop-add and before the midpoint, they receive a W (unless they already have 6 withdrawals); if they withdraw after the midpoint, they will automatically receive a WF. After 6 withdrawals, withdrawal at any point in the course results in an automatic F.
While Voluntary Withdrawals are the most common, GSU policy also permits Involuntary Withdrawals, Emergency Withdrawals, Military Withdrawals, and Non-Academic Withdrawals, and explains when and how students can initiate a withdrawal. You are responsible for understanding and adhering to the GSU Revision of Class Schedule (Add/Drop and Withdrawal) policy explained in Section 1332 of the Undergraduate Catalog.

9.4 Campus safety APP (Livesafe Mobile App)

Georgia State University values the safety of all university community members. To promote campus safety, the university is providing the LiveSafe app free for all students, faculty, and staff. This app provides a quick, convenient, and discrete way to communicate with the GSU police. I strongly recommend that you download the app from either the Apple App Store or Google Play. You can sign-up for Panther Alerts and learn more about LiveSafe by visiting the GSU LiveSafe webpage.

9.5 Campus police numbers

Please make sure you have these campus police numbers in your phone

- For emergencies call 404-413-3333
- For non-emergencies and to request a safety escort call 404-413-2100
- If you are hearing impaired call 404-413-3203

9.6 Family educational rights privacy act (FERPA)

In keeping with USG and university policy, this course website will make every effort to maintain the privacy and accuracy of your personal information. Specifically, unless otherwise noted, it will not actively share personal information gathered from the site with anyone except university employees whose responsibilities require access to said records. However, some information collected from the site may be subject to the Georgia Open Records Act. This means that while we do not actively share information, in some cases we may be compelled by law to release information gathered from the site. Also, the site will be managed in compliance with the Family Educational Rights and Privacy Act (FERPA), which prohibits the release of education records without student permission. For more details on FERPA, go here.

9.7 Course assessment

"Your constructive assessment of this course plays an indispensable role in shaping education at Georgia State. Upon completing the course, please take the time to fill out the online course evaluation."

9.8 Academic honesty

The GSU Policy on Academic Honesty applies to all your GSU courses, including LGLS 3610. The policy defines and provides examples of several types of academic dishonesty, including plagiarism, cheating on exams, unauthorized collaboration, falsification, and multiple submissions. In addition, the policy outlines the possible consequences for academic dishonesty, including failing the plagiarized assignment, failing the course, an annotation on your transcript, and even expulsion from the university.

9.9 Student code of conduct (2022-2023)

The Student Code of Conduct addresses a number of issues, including general student conduct that is prohibited by the university, disruptive student conduct, the university's non-discrimination policy, and the sexual misconduct policy. For more information on any of these policies refer to the Code of Conduct.