MSA 8010: Data Programming (Section 002)
Syllabus for Summer 2023

Instructor: Xiangshi Yin*

June 14, 2023

Contents

---

*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	TBD
Email	NA

1.3 Lectures

We meet on every Monday and Wednesday evening at 7:40 PM (Eastern Time).

Days	Monday and Wednesday
Time	7:40 PM - 10:10 PM (Eastern Time)
Room	(Online) Webex@iCollege

To attend the online class, you need to:

- Go to the class home page on iCollege at
  `https://gastate.view.usg.edu/d2l/home/2822689`

- Click Webex tab→Click Virtual Meetings →Choose the corresponding class link and join.

- You could also click Content   Course Schedule →Choose the corresponding class link and join

1.4 Office Hours

Days	Monday and Wednesday
Type	By Appointment
Time	7:00 PM - 7:30 PM (Eastern Time)
Room	(Online) Webex@iCollege

There are two 30 minutes sessions available every Monday and Wednesday before our regular classes. To book the time on calendar, you need to:

- Go to the class home page on iCollege at
  `https://gastate.view.usg.edu/d2l/home/2822689`

- Click Webex tab→Click Office Hours →Choose the available time slot and click Book .

- After the meeting is booked, you'll receive Webex online meeting instructions in your school email address, and you can also add the meeting to your calendar so that you don't miss it.

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 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	2023-06-05	Python Crash Course (p1)	link
02	2023-06-07	Python Crash Course (p2)	link
03	2023-06-12	Python Crash Course (p3)	link
04	2023-06-14	Numpy and Linear Algebra	link
05	2023-06-19	Juneteenth (**No Classes**)	(online)
06	2023-06-21	Pandas and Data Manipulation	(online)
07	2023-06-26	Python and Statistics	(online)
08	2023-06-28	Data Visualization	(online)
09	2023-07-03	Introduction to Machine Learning (p1)	ML Chapter 2,8
10	2023-07-05	Introduction to Machine Learning (p2)	ML Chapter 3
11	2023-07-10	Information Based Learning	ML Chapter 4
12	2023-07-12	Similarity Based Learning	ML Chapter 5
13	2023-07-17	Probability Based Learning	ML Chapter 6
14	2023-07-19	Error Based Learning	ML Chapter 7
15	2023-07-24	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

- **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

- **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

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.

- 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 4 quizzes, 4-5 home assignments, and 1 final group project over the whole semster. Detailed due dates are listed below.

Date	Note	Quizz	Homework	Quizz/Homework Due(EST)
2023-06-05	First Class			N/A
2023-06-08		Y		2023-06-13 23:59
2023-06-14			Y	2023-06-20 19:29
2023-06-19	Juneteenth (**No Classes**)			N/A
2023-06-21		Y		2023-06-28 19:29
2023-06-26	Team assigned		Y	2023-07-03 19:29
2023-06-28	Team finalized	Y		2023-07-05 19:29
2023-07-03	Suggested project topics		Y	2023-07-10 19:29
2023-07-05		Y		2023-07-12 19:29
2023-07-10	Project topics finalized		Y	2023-07-17 19:29
2023-07-12				N/A
2023-07-17				N/A
2023-07-19				N/A
2023-07-24	Project Presentation			2023-07-28 00:00

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[1] 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):

- 2023-06-26: Project teams will be assigned[2]
- 2023-06-28: Project team will be finalized[3]
- 2023-07-03: The instructor will release a list of suggested project topics
- 2023-07-10: Team leads need to submit their teams' choice on final project topics
- 2023-07-24: Final project presentation
- 2023-07-27: 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:

---

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

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 Student Behavior

Behavior in class should be professional at all times. People must treat each other with dignity and respect in order for scholarship to thrive. Behaviors that are disruptive to learning will not be tolerated and may be referred to the Office of the Dean of Students for disciplinary action.

9.1 Discrimination and harassment

Discrimination and/or harassment will not be tolerated in the classroom. In most cases, discrimination and/or harassment violates Federal and State laws and/or University Policies and Regulations. Intentional discrimination and/or harassment will be referred to the Affirmative Action Office and dealt with in accordance with the appropriate rules and regulations. Unintentional discrimination and/or harassment is just as damaging to the offended party. But, it usually results from people not understanding the impact of their remarks or actions on others, or insensitivity to the feelings of others. We must all strive to work together to create a positive learning environment. This means that each individual should be sensitive to the feelings of others, and tolerant of the remarks and actions of others. If you find the remarks and actions of another individual to be offensive, please bring it to their attention. If you believe those remarks and actions constitute intentional discrimination and/or harassment, please bring it to my attention.

9.2 Official department class policies

- Students are expected to attend all classes and group meetings, except when precluded by emergencies, religious holidays, or bona fide extenuating circumstances.

- Students who, for non-academic reasons beyond their control, are unable to meet the full requirements of the course should notify the instructor, by email, as soon as this is known and prior to the class meeting. Incompletes may be given if a student has ONE AND ONLY ONE outstanding assignment.

- A "W" grade will be assigned if a student withdraws before mid-semester if (and only if) he/she has maintained a passing grade up to the point of withdrawal. Withdrawals after the mid-semester date will result in a grade of "WF". See the GSU catalog or registrar's office for details.

- Spirited class participation is encouraged and informed discussion in class is expected. This requires completing readings and assignments before class.

- All exams and individual assignments are to be completed by the student alone with no help from any other person.

- Collaboration within groups is encouraged for project work. However, collaboration between project groups will be considered cheating.

- Copying work from the Internet without a proper reference is considered plagiarism and subject to disciplinary action as delineated in the GSU Student Handbook.

- Any non-authorized collaboration will be considered cheating and the student(s) involved will have an Academic Dishonesty charge completed by the instructor and placed on file in the Dean's office and the CIS Department. All instructors regardless of the type of assignment will apply this Academic Dishonesty policy equally to all students. Abstracted from GSU's Student Handbook Student Code of Conduct "Policy on Academic Honesty and Procedures for Resolving Matters of Academic Honesty" (`https://codeofconduct.gsu.edu/`)

As members of the academic community, students are expected to recognize and uphold standards of intellectual and academic integrity. The University assumes as a basic and minimum standard of conduct in academic matters that students be honest and that they submit for credit only the

products of their own efforts. Both the ideals of scholarship and the need for fairness require that all dishonest work be rejected as a basis for academic credit. They also require that students refrain from any and all forms of dishonorable or unethical conduct related to their academic work.

Students are expected to discuss with faculty the expectations regarding course assignments and standards of conduct. Here are some examples and definitions that clarify the standards by which academic honesty and academically honorable conduct are judged at GSU.

Plagiarism. Plagiarism is presenting another person's work as one's own. Plagiarism includes any paraphrasing or summarizing of the works of another person without acknowledgment, including the submitting of another student's work as one's own. Plagiarism frequently involves a failure to acknowledge in the text, notes, or footnotes the quotation of the paragraphs, sentences, or even a few phrases written or spoken by someone else. The submission of research or completed papers or projects by someone else is plagiarism, as is the unacknowledged use of research sources gathered by someone else when that use is specifically forbidden by the faculty member. Failure to indicate the extent and nature of one's reliance on other sources is also a form of plagiarism. Any work, in whole or part, taken from the Internet or other computer based resource without properly referencing the source (for example, the URL) is considered plagiarism. A complete reference is required in order that all parties may locate and view the original source. Finally, there may be forms of plagiarism that are unique to an individual discipline or course, examples of which should be provided in advance by the faculty member. The student is responsible for understanding the legitimate use of sources, the appropriate ways of acknowledging academic, scholarly or creative indebtedness, and the consequences of violating this responsibility.

Cheating on Examinations. Cheating on examinations involves giving or receiving unauthorized help before, during, or after an examination. Examples of unauthorized help include the use of notes, texts, or "crib sheets" during an examination (unless specifically approved by the faculty member), or sharing information with another student during an examination (unless specifically approved by the faculty member). Other examples include intentionally allowing another student to view one's own examination and collaboration before or after an examination if such collaboration is specifically forbidden by the faculty member. Unauthorized Collaboration. Submission for academic credit of a work product, or a part thereof, represented as its being one's own effort, which has been developed in substantial collaboration with another person or source or with a computer-based resource is a violation of academic honesty. It is also a violation of academic honesty knowingly to provide such assistance. Collaborative work specifically authorized by a faculty member is allowed.

Falsification. It is a violation of academic honesty to misrepresent material or fabricate information in an academic exercise, assignment or proceeding (e.g., false or misleading citation of sources, the falsification of the results of experiments or of computer data, false or misleading information in an academic context in order to gain an unfair advantage).

Multiple Submissions. It is a violation of academic honesty to submit substantial portions of the same work for credit more than once without the explicit consent of the faculty member(s) to whom the material is submitted for additional credit. In cases in which there is a natural development of research or knowledge in a sequence of courses, use of prior work may be desirable, even required; however the student is responsible for indicating in writing, as a part of such use, that the current work submitted for credit is cumulative in nature.