Northwestern | McCORMICK SCHOOL OF ENGINEERING Master of Science in Machine Learning and Data Science

MLDS 422: Python & Other Data Science Programming

Fall 2023 Quarter

Instructor: Alice Zhao

Contact: alicezhao2013@u.northwestern.edu

Class Meets: Thursdays 1:00PM – 4:00PM (Krebs classroom)

Office Hours: Thursdays 11:00AM – 12:00PM (drop in or by appointment, MEC 431)

Lab & Project Grading TA: Sungsoo Lim sungsoolim2024@u.northwestern.edu

TA Lab Hours: Wednesdays 11:00AM – 12:00PM (Krebs classroom)

TA Office Hours: Mondays 2:30PM – 3:30PM (remote by appointment)

Homework Grading TA: Holly Wang jiayiwang2020@u.northwestern.edu

I. **Course description:** The main objective of this course is to introduce students to programming languages for data science, mainly focusing on Python, with additional overviews of Julia, Scala and Rust. Students will be introduced to common libraries, approaches and workflows used by data scientists in Python. In addition, students will be exposed to broader programming principles and its applications within Python.

II. **Recommended text and/or other materials:**
   a. Learn Python the Hard Way – Zed Shaw – Addison-Wesley Professional
   b. Python for Data Analysis – Wes McKinney – O'Reilly
   c. Learning Python, Powerful Object-Oriented Programming – Mark Lutz – O'Reilly

III. **Course outcomes:** After finishing the course, students should be comfortable:
   a. Knowing the fundamentals of Python, including writing working, well-organized and high-quality Python code
   b. Using both base Python and imported libraries to solve data science problems
   c. Reading code in other coding languages and adjusting to new development environments

IV. Tentative Course Outline

Week 1 *Homework#1 Assigned*	- Tools: Anaconda, Jupyter Notebook - Lists, dictionaries, loops, list comprehensions, functions, etc. - Pythonic coding, virtual environments
Week 2 *Homework#2 Assigned*	- Lambda functions, iterators, generators, recursion, etc. - Importing libraries, numpy, vectorization - Timeit, complexity, search & sort algorithms
Week 3 *Homework#3 Assigned*	- OOP: objects, classes - Pandas: DataFrames, data manipulation, chaining methods
Week 4 *Project#1 Assigned*	- More Pandas: data types, data cleaning, pickling, exporting - Data visualization: matplotlib, seaborn
Weeks 5	- Regular expressions - Web scraping: BeautifulSoup, selenium - Data structures: binary trees, hash tables, etc.
Week 6 *Homework#4 Assigned*	- Integration with databases, APIs - Threading, multi-threading
Week 7 *Homework#5 Assigned*	- Feature engineering: transformations, one-hot encoding, etc. - Data science workflow with pandas, feature engineering, sklearn
Week 8: Julia	- Basic programming, differences compared to Python, pros and cons - Data science with Julia
Week 9: Scala *Project#2 Assigned*	- Basic programming, differences compared to Python, pros and cons - Data science with Scala
Week 10: Rust	- Basic programming, differences compared to Python, pros and cons - Data science with Rust

V. Grading: Your grades will be based on:

- Five homework assignments (12% each – total 60%)
- Two projects (20% each – 40%)

Please note: The distribution given above is approximate and may be subject to some very minor changes. The firm policy will be announced during the last week of classes.

Awareness, Academic Responsibilities and Closing Remarks:

Please be advised that it is each student's *individual responsibility* to keep him/herself up-to-date with the announcements *made in class*, *distributed via email*, or *otherwise posted*.

Programming projects will most likely be done in teams, however, homework is indicated as *individual* – although you are encouraged to always discuss class-related issued with your classmates – it is your responsibility to ensure that the work is done individually. Example: you are encouraged to discuss the algorithmic aspects of solving a particular programming assignment, and even the high-level design approach. However, the source code that you will submit must be typed individually in its entirety.

The policies for cheating are well-defined and there will be no exceptions made for any excuse whatsoever – if caught cheating (both in terms of borrowing someone else's code, as well as allowing someone to borrow your code), you will automatically fail the class. In addition, notwithstanding our willingness to be understanding for the students' commitments and time-constraints, please do not attempt to obtain an incomplete grade for the course, based solely on your poor performance – it is against the University regulations.

Lastly, please note that your grade is based on the programming assignment. Hence, you really need to keep yourself up-to-date with the material lectured and start working on the programs as early as possible. You should not allow yourself to fall behind with the topics, as the new ones will be building incrementally upon the older ones, and it will be very hard to catch up. Plan your time wisely.

Academic Integrity Statement

Students in this course are required to comply with the policies found in the booklet, "Academic Integrity at Northwestern University: A Basic Guide". All papers submitted for credit in this course must be submitted electronically unless otherwise instructed by the professor. Your written work may be tested for plagiarized content. For details regarding academic integrity at Northwestern or to download the guide, visit: https://www.northwestern.edu/provost/policies- procedures/academic-integrity/index.html

Accessibility Statement

Include the following statement on all course syllabi.

Northwestern University is committed to providing the most accessible learning environment as possible for students with disabilities. Should you anticipate or experience disability-related barriers in the academic setting, please contact AccessibleNU to move forward with the university's established accommodation process (e: accessiblenu@northwestern.edu; p: 847-467-5530). If you already have established accommodations with AccessibleNU, please let me know as soon as possible, preferably within the first two weeks of the term, so we can work together to implement your disability accommodations. Disability information, including academic accommodations, is confidential under the Family Educational Rights and Privacy Act.

COVID-19 Classroom Expectations Statement

Students, faculty and staff must comply with University expectations regarding appropriate classroom behavior, including those outlined below and in the COVID-19 Expectations for Students. With respect to classroom procedures, this includes:

- Policies regarding masking, social distancing and other public health measures evolve as the situation changes. Students are responsible for understanding and complying with current University, state and city requirements.
- In some classes, masking and/or social distancing may be required as a result of an Americans with Disabilities Act (ADA) accommodation for the instructor or a student in the class even when not generally required on campus. In such cases, the instructor will notify the class.

If a student fails to comply with the COVID-19 Expectations for Students or other University expectations related to COVID-19, the instructor may ask the student to leave the class. The instructor is asked to report the incident to the Office of Community Standards for additional follow-up.

Exceptions to Class Modality

Maintaining the health of the community remains our priority. If you are experiencing any symptoms of COVID do not attend class. Follow the steps outlined on this site for testing, isolation and reporting a positive case. Next, contact your instructor as soon as possible to arrange to complete coursework.

Students who experience other personal emergencies should contact the instructor as soon as possible to arrange to complete coursework.

Should public health recommendations prevent in-person class from being held on a given day, the instructor or the university will notify students.

Prohibition of Recording Classes by Students

Unauthorized student recording of classroom or other academic activities (including advising sessions or office hours) is prohibited. Unauthorized recording is unethical and may also be a violation of University policy and state law. Students requesting the use of assistive technology as an accommodation should contact AccessibleNU. Unauthorized use of classroom recordings – including distributing or posting them – is also prohibited. Under the University's Copyright Policy, faculty own the copyright to instructional materials – including those resources created specifically for the purposes of instruction, such as syllabi, lectures and lecture notes, and presentations. Students cannot copy, reproduce, display, or distribute these materials. Students who engage in unauthorized recording, unauthorized use of a recording, or unauthorized distribution of instructional materials will be referred to the appropriate University office for follow-up.

Support for Wellness and Mental Health

Northwestern University is committed to supporting the wellness of our students. Student Affairs has multiple resources to support student wellness and mental health. If you are feeling distressed or overwhelmed, please reach out for help. Students can access confidential resources through the Counseling and Psychological Services (CAPS), Religious and Spiritual Life (RSL) and the Center for Awareness, Response and Education (CARE). Additional information on all of the resources mentioned above can be found here:

https://www.northwestern.edu/counseling/

https://www.northwestern.edu/religious-life/

https://www.northwestern.edu/care/