

**INFO 8750 (CRN 65886):
Advance Programming for Data Mining
University of Georgia – College of Engineering
Spring 2023**

Instructor

Paul Zhong-Ru Xie

Room 145 Boyd Graduate Studies Research Center
paulxie@uga.edu

706-542-9923 (office)

Office Hours: After class and by appointment

Class Times & Venue

Tue./Thr. 12:45-14:00

Driftmier 1460

Offered

Credits

Level

**Weekly Class Meeting
Pattern**

Spring

3

Graduate

3 hour lecture & lab classes

Course Pre- or Co-requisites

Permission of instructor

eLearning Commons and Email Requirement

This class will utilize eLC (elc.uga.edu) for assignments and various other communications. Students are required to keep up with new postings on eLC.

Email (your UGA email address) may be used to give assignments and **WILL** be used to make announcements.

**Students are required to complete any assignments given using email and eLC, so
CHECK BOTH OF THESE DAILY!**

Overview of Course

The main goals of this course is to for students to learn programming skills and learn how to apply them to “**Big Data**”. This interdisciplinary course will guide students in studying the process of discovering new knowledge from “Big Data” with applications of data retrieval, data processing, data analysis, modeling, and visualization. Python, which is a simple and

popular programming language, will be taught mainly. No prerequisite programming experience is needed for taking this course. Topics include real world examples, e.g. processing and analyzing large-scale data. The idea of “Big Data” will be emphasized to help students with their coding skills to discovering new knowledge in natural or social sciences and solving some real-world problems.

Course Format

Class meetings will consist of lectures introducing computational tools and algorithms, group discussion, computational lab for student to practice coding, search online resources and work with online databases and servers to finish different tasks. Students must actively participate in discussions and integrate all the knowledge and skills learned to finish the assigned project. *The course syllabus is a general plan for the course; deviation announced by the instructor may be necessary.*

Topical Outline

Topics will include:

- Programming in Python
- Online databases
- Automatic data retrieval
- Data processing
- Dynamic programming
- Programming for kinetic or other models
- Simulation models
- Programming for data analysis
- Programming for basic statistical analysis
- Clustering and classification
- Prediction models

Grade Determination

This course will consist of in-class exercises, assignments, a term paper of a final project. Students will be graded on their participation, professional attitude, knowledge and understanding of programming concepts and data, and communication skills.

The following is the grade distribution for this course:

Grading System:	Assignments**	30-40%	Grades:	90-100 A
	Paper review	10-20%		80-89 B
	<u>Project Report</u>	<u>50-60%</u>		70-79 C

Total 100%

65-69 D

< 65 F

***Note: Some assignments may be worth more than others.*

I don't GIVE grades. YOU must EARN them. We will do what we can to help you earn good grades, so if you have a problem, let instructor know and we will try to help you!!!

How to Make the Most of the Classroom

1. Finish and deliver each assignment on time based on what you have learned on the class.
2. When you do your assignments and assigned project, try to apply the coding concepts, logic, and optimize the parameters and procedures, understand the strengths and weaknesses of different algorithms and explain why you use a tool or server to address a problem.
3. Pair with your classmates with different background and expertise for the project. This will help you deliver a complete project report/presentation that covers various aspects and gain knowledge/skills in a field that is different from yours.

Course Work

In-class exercises, discussions, assignments and a final project will be used to reinforce material presented in lectures.

In-Class Exercises: In-class exercises, including coding, database searching, and retrieving and processing data, will provide students with opportunities to practice and reinforce what they learned in the course, to analyze and familiarize themselves with data, and to gain the skills to retrieve and process enormous amount of data online. These exercises will assist each student in finishing the assignments and final project.

Assignments: Assignments will be given primarily to assess students' understanding of different programming operations and their application in data collection, processing, and analysis, and students' efforts to try different approaches, compare the differences among them, optimize the efficiency and accuracy by good reasoning, and identify a most suitable algorithm(s) or programming strategy(ies) for a problem. For each topic, a corresponding assignment will be given after the first lecture of the topic. For example, a protein structure (PDB) file will be assigned to a student after the topic "processing structural data of protein" has been taught. Students will be asked to write a script to parse the PDB file, generate the amino acid sequence of the protein, and compute the percentage of 20 amino acids in this protein.

Assignments are typically due on Mondays *before* midnight (see schedule). Late submission will receive 50% credit.

Final Project: The final project will provide students with an opportunity to integrate all the knowledge and skills they have learned in this course to finish a realistic task -- to create an automatic program to distinguish two groups of proteins with different biological functions. Programming proficiency, reasoning, creativity, prediction accuracy and efficiency, understanding of data, and written communications will be assessed to determine the grade of the final project. The final project will be discussed further when it is assigned in the middle of the semester.

Engineering Grading Policy Regarding Communication Skills

Thirty percent (30%) of the grade on all written assignments and final paper will be based on quality of communication. Spelling, grammar, punctuation, and clarity of writing are evidence of written communication quality. Enunciation, voice projection, clarity and logical order of the presentation and effective use of visual aids are evidence of oral communication quality.

Personal issues (medical and otherwise) that may lead to late or un-submitted work (including assignments and the project report) will be handled on a case by case basis.

It is the YOUR responsibility to seek assistance from the instructor if you are experiencing trouble understanding the materials presented in this course. If you find that you need assistance outside of class, please schedule a meeting with me.

This course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary.

Course Communication

Typically, the instructor will communicate by email. These emails might include supplemental material, assignment reminders, and general course information. Students must provide a valid email address through elc and it is the student's responsibility to check this email. Students are encouraged to send the instructor emails if assignments are unclear or if he/she would like to schedule an office appointment.

Academic Honesty

As a University of Georgia student, you have agreed to abide by the University's academic honesty policy, "A Culture of Honesty," and the Student Honor Code. All academic work must meet the standards described in "A Culture of Honesty" found at: <https://ovpi.uga.edu/academic-honesty/academic-honesty-policy>. Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation. Questions related to course assignments and the academic honesty policy should be directed to the instructor.

Engineering Professionalism Policy

Engineers make great contributions to society. Engineering is a very satisfying profession that provides many rewards but is demanding and requires hard work. The engineering profession is governed by a code of ethics. The following link will take you to the National Society of Professional Engineers, Engineering Code of Ethics website. <http://www.nspe.org/resources/ethics/code-ethics>. Engineering faculty at UGA expect students to act in a professional manner at all times and develop the work ethic required for a

successful engineering career. Engineering students at UGA are responsible for maintaining the highest standards of professionalism and professional practice.

Out of the Classroom Meetings with Instructors

Out of the classroom meetings with me can be arranged as needed. In order to develop professional skills, it is preferred that students contact the instructor in person or by email to set a date and time for the meeting. If you make an appointment and must cancel, it is expected that you contact the instructor in a reasonable amount of time. Unscheduled meetings to discuss course material, grading or other issues directly related to the course are discouraged, unless during specific drop-in office hours. If you are having trouble, PLEASE schedule a meeting in a timely manner.