CSCI 3360 Data Science I
Syllabus

Spring 2024

Instructor: Prof. Jin Sun
4 Credit Hours

Catalog Description: This class is designed as an introductory study of the theory and practice of data science. Topics covered include fundamentals of data science, practical libraries to handle data, data collection and cleaning, data visualization and analysis, learning algorithms for classification and regression, unsupervised learning, validation metrics, applications in computer vision, natural language processing, and recommendation systems.

Class webpage: https://jinsungit.github.io/teaching/Spring2024/

Prerequisites: CSCI 1302, 2610, 2720. Students should have a solid understanding of calculus and linear algebra.

Class Location and Times:

Tue & Thu	2:20 pm - 3:35 pm	Miller Plant Science 1102
Wed	3:00 pm – 3:50 pm	Conner Hall 0210

Textbooks:
- An introduction to statistical learning (https://www.statlearning.com/)
- Python data science handbook (https://jakevdp.github.io/PythonDataScienceHandbook/)

Student Outcomes:
1. Demonstrate understanding of data science pipeline fundamentals.
2. Familiar with relevant data science libraries and software packages.
3. Ability to formulate a learning problem from data.
4. Ability to evaluate a learning model.
5. Gain experience deploying learning models in computer vision, natural language processing, and other application domains.

Instructor Contact:
Name: Prof. Jin Sun
Office Hours: Thursdays 4-5pm or by appointment
Office: 804 Boyd
Email: jinsun@uga.edu

Teaching Assistant:
TBD

Evaluation and Grading: The final course grade will be weighted as follows:

Quiz:	10%
Homework:	40%
Midterm exam:	10%
Final exam:	15%
Project:	25%

Homework assignments: Your homework submission should be in PDF format. You are encouraged to use LaTeX (for online editing, use Overleaf). Cite any references used (including books, online webpages, and code). Your homework should be done by yourself, not as a group.

The PDF file should contain all essential text, equations, figures, code, and program outputs. Attach your code as the appendix.

Exams: You will take an in-class midterm and final exams to test your knowledge on the essential concepts covered in the class.

Project: You will work in a team on a course project. Each team should have 2-3 members. You are encouraged to design the project to solve a real-world application using data science. Feel free to use any programming language or software packages of your choice. The schedule for the project is as follows:

1. **Project Proposal (5%)**: The project proposal should clearly state what your team plan to do. It should be four pages long (not including references). It should contain a timeline. You should list the questions the project will address and that will be discussed in the report. You should list what software you will be using or will build upon. Describe the datasets you will use and how will you know if the project is successful. Describe the hypotheses you will test and the related work. You should be able to reuse much of the text for the final report.

2. **Project Milestone (5%)**: You can re-use the project proposal for this report but expand it with additional content. You should talk about preliminary results and/or other measurable items listed in the proposal.

3. **Project Report and Presentation (15%)**: The final report contains a complete description of the project: what you have done and what the result looks like. It should be about six to eight pages long (not including references). You are encouraged to format it in [CVPR](#) format. We will have presentation sessions for all projects during the last week of the class. Make sure every member in your team participate in the presentation.

Late policy:

All homework and project reports are expected to be submitted to eLC by the due date. Late submission will receive 10% (of total scores for the assignment) penalty for each day passed.

Course Topics:

Data science pipeline, kaggle, python basics and development setup, data collection and processing, data visualization, probabilistic learning, regression, classification, unsupervised learning,

trees and boosting, validation and metrics, vision, language, recommendation systems, real-world tools and development.

UGA Well-being Resources

UGA Well-being Resources promote student success by cultivating a culture that supports a more active, healthy, and engaged student community.

Anyone needing assistance is encouraged to contact Student Care & Outreach (SCO) in the Division of Student Affairs at 706-542-8479 or visit sco.uga.edu. Student Care & Outreach helps students navigate difficult circumstances by connecting them with the most appropriate resources or services. They also administer the Embark@UGA program which supports students experiencing, or who have experienced, homelessness, foster care, or housing insecurity.

UGA provides both clinical and non-clinical options to support student well-being and mental health, any time, any place. Whether on campus, or studying from home or abroad, UGA Well-being Resources are here to help.

• Well-being Resources: well-being.uga.edu

• Student Care and Outreach: sco.uga.edu

• University Health Center: healthcenter.uga.edu

• Counseling and Psychiatric Services: caps.uga.edu or CAPS 24/7 crisis support at 706-542-2273

• Health Promotion/ Fontaine Center: healthpromotion.uga.edu

• Disability Resource Center and Testing Services: drc.uga.edu

Additional information, including free digital well-being resources, can be accessed through the UGA app or by visiting https://well-being.uga.edu.

School of Computing
Policy Statement on Academic Honesty

The Computer Science Department recognizes honesty and integrity as necessary to the academic function of the University. Therefore all students are reminded that the CS faculty requires compliance with the conduct regulations found in the University of Georgia Student Handbook. Academic honesty means that any work you submit is your own work.

Common forms of academic dishonesty against which students should guard are:

1. Copying from another student's test paper or laboratory report, or allowing another student to copy from you;
2. Fabricating data (computer, statistical) for an assignment;
3. Helping another student to write a laboratory report or computer software code that the student will present as his own work, or accepting such help and presenting the work as your own;
4. Turning in material from a public source such as a book or the Internet as your own work.

Three steps to help prevent academic dishonesty are:

1. Familiarize yourself with the regulations.
2. If you have any doubt about what constitutes academic dishonesty, ask your instructor or a staff member at the Office of Judicial Programs.
3. Refuse to assist students who want to cheat.

All faculty, staff and students are encouraged to report all suspected cases of academic dishonesty. All cases of suspected academic dishonesty (cheating) will be referred to the Office of the Vice President for Instruction. Penalties imposed by the Office of Judicial Programs may include a failing grade in the course and a notation on the student's transcript. Repeated violations are punishable by expulsion from the University. For further information please refer to the UGA Code of Conduct, available at the URL below.

https://honesty.uga.edu/Academic-Honesty-Policy/