CSE 2600: Introduction to Data Science and Engineering

Syllabus – Summer 2025

Excluding materials for purchase, syllabus information may be subject to change. The most up-to-date syllabus is located within the course in HuskyCT.

Course and Instructor Information

Course Title: Introduction to Data Science and Engineering
Credits: 3
Mode of Instruction: Online/Asynchronous
Prerequisites: CSE 1010. Not open for credit to students who have passed CSE 4095 when offered as "Introduction to Data Science and Engineering." May not be taken out of sequence after passing CSE 4820 or CSE 5819.
Professor: Matt Lamoureux

Email: matthew.lamoureux@uconn.edu (preferred method of contact)
Office Hours/Availability: TBD. Most emails should receive reply in 24 hours.

Course Materials

Required course materials should be obtained before the first day of class.

Required Materials:
James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning with Applications in Python*. Springer. Available online at https://www.statlearning.com

ISLP Python Environment: https://intro-stat-learning.github.io/ISLP/installation.html

The University has set minimum device requirements for all students.

Optional Materials:
Grus, J. (2019). *Data Science from Scratch: First Principles with Python*, 2nd ed. O'Reilly Media.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann.

Additional instructional materials and links to resources are available from within the HuskyCT course.

Course Description

Introduction to a broad selection of challenges and methodologies in working with big data. Topics to be covered include fundamental data science lifecycle topics such as data acquisition, management, integration, visualization, modeling, analysis, prediction, as well as data security, data privacy, and ethics.

The goal of this course is to prepare students for **career readiness** and/or **research** in data science.

Course Objectives

By the end of the semester, students should be able to:

1. (the foundation) Explain the role and importance of data science in various domains.
2. (the engineering) Preprocess large datasets via common manipulation techniques in Python.

3. (the science) Apply statistical methods and machine learning algorithms to derive insights.
4. (the communication) Communicate analysis effectively via data visualizations and technical explanations.
5. (the ethics) Discuss the ethical considerations in practicing data science and engineering.

<div align="center">

Course Outline

</div>

A. **Module 1 - Preliminaries**
   a. Intro
   b. Data Objects and Attributes & Measures of Central Tendency
      Activity: Review of Data Manipulations in Python
   c. Data Visualizations
      Activity: Matplotlib
   d. Foundations of Modeling (Ch. 2)
   e. Simple Linear Regression (Ch. 3)
      Activity: Navigating File Systems
   f. HOMEWORK 1

B. **Module 2 – Linear Regression**
   a. Git: Global information tracker; **Guest speaker**: Maggie
   b. Data preprocessing (binning & normalization)
      Activity: Fitting lines
   c. Multiple Linear Regression (Ch. 3)
   d. Interactions
      Activity: Best linear regression model
   e. Probability & LDA (Ch. 4)
   f. Classification via Logistic Regression (Ch. 4)
      Activity: Best classification model
   g. Group Project Milestone: Form groups
   h. HOMEWORK 2

C. **Module 3 – Classification & Clustering**
   a. Resampling (Ch. 5)
   b. Random Forests (Ch. 8)
      Activity: Hyperparameter Tuning
   c. Boosting (Ch. 8)
   d. Model Comparison & Visualizations
      Activity: RF vs GBM
   e. Group Project Milestone: Select data
   f. HOMEWORK 3
   g. EXAM

D. **Module 4 – Model Aggregation**
   a. Dimensionality Reduction: PCA (Ch. 12)
   b. Clustering (Ch. 12)
      Activity: PCA and Clustering
   c. Databases & SQL
      Activity: SQL Detective site
   d. **Guest speaker**: TBD
   e. Milestone for Group Project: Submit abstract
   f. HOMEWORK 4

E. **Module 5 – Additional Topics**
   a. SQL
   b. Ethics
   c. HOMEWORK 5
   d. P**roject Presentations**

Summary of Course Grading:

Course Components	Weight
Participation	10%
Homework	20%
Exam	30%
Final Project Report	20%
Final Project Presentation	20%

Participation
- Because this course is fast-paced, timely viewings (and re-viewings) of lectures will be critical to achieving learning outcomes. Take notes that can help with future work and post questions as you encounter new and/or confusing material.
- Most classes will involve some lab-style coding work done in real time. These activities are meant to better solidify the concepts seen in lecture videos.
- Each module contains a discussion board that will require you to post your own original ideas as a solution to a problem, and to interact with other students by posting responses.
- There will be weekly assessments submitted on HuskyCT to help us (student and instructor) learn what concepts have or haven't been mastered. These are not meant to be intimidating, and should not take more than 10 minutes to complete. These will be due most weekdays at midnight, and you can miss 2 without penalty—if you don't send me any emails requesting extensions.

Homework
- There will be 4 assignments that must be submitted via one PDF on HuskyCT.
- Deadlines will be available on HuskyCT and generally at midnight. You can submit a late assignment only if: (1) grading has not yet begun, (2) the solutions have not yet been posted, and (3) you have not emailed me requesting an extension. To be safe, start each assignment early and submit whatever work is completed by the deadline.
- It is **highly** recommended that students form study groups and work together on homework assignments. Please be sure to include the names of your study group members on your homework submission as a form of acknowledging their contribution.

Exam
- Will be conducted on HuskyCT with Lockdown Browser. Expect the exam to take about an hour to complete.
- Multiple choice questions (30-40%) based on the information from the lecture slides will assess fact recognition, important definitions, etc. Some of these questions may be directly taken from the concept check quizzes or the hints provided in lecture videos.
- Free response questions (60-70%) will allow you to apply methods from class and are likely to be formatted similarly to homework questions. There will not be live coding during the exam.
- Requests for accommodations (through CSD) are welcomed and should be arranged at the beginning of the semester, or at least one week before an exam.

Final Project
- You will collaborate in pairs on a topic of your choice. Once groups are finalized, you will need to submit a short abstract describing your project proposal on HuskyCT.
- Both the live presentation, to be scheduled at the end of the class, and the written report will allow you to demonstrate your ability to communicate data analysis effectively.
- Further details can be found on HuskyCT.

Grading Scale:

Information on grades and grading can be found on the Registrar's site and in the catalog:

-
-

Grade	Letter Grade	Points
Excellent	A	**93-100**
	A-	**90-92**
Very Good	B+	**87-89**
Good	B	**83-86**
	B-	**80-82**
	C+	**77-79**
Average	C	**73-76**
Fair	C-	**70-72**
Poor	D+	**67-69**
	D	**63-66**
Merely Passing	D-	**60-62**
Failure	F	**00-59**

Due Dates and Late Policy

All course due dates are identified in HuskyCT. Deadlines are based on Eastern Time; if you are in a different time zone, please adjust your submittal times accordingly. *The instructor reserves the right to change dates accordingly as the semester progresses. All changes will be communicated in an appropriate manner.*

No late submissions of quizzes will be accepted. For homework assignments, you may submit a late assignment only if: (1) grading has not yet begun, (2) the solutions have not yet been posted, and (3) you have not emailed me requesting an extension. To be safe, start each assignment early and submit whatever work is completed by the deadline.

Feedback and Grades

I will make every effort to provide feedback and grades in the comments of assignments. To keep track of your performance in the course, refer to Gradebook in HuskyCT.

Weekly Time Commitment

You should expect to dedicate 25-30 hours a week to this course. This expectation is based on the various course activities, assignments, and assessments and the University of Connecticut's policy regarding credit hours. More information related to hours per week per credit can be accessed at the [Online Student website](#).

Student Authentication and Verification
The University of Connecticut is required to verify the identity of students who participate in online courses and to establish that students who register in an online course are the same students who participate in, complete the course activities and assessments, and receive academic credit. Verification and authentication of student identity in this course will include:
1. Secure access to the learning management system using your unique UConn NetID and password.

2. Recorded video submitted by student and live video during final presentation. In both cases, identity is verified via comparison to photo found on Student Administration System.

Students who do not complete the above required authentication steps may be denied access to the course and given an incomplete. Students could lose credit if the identity of the enrolled student completing course activities and assessments cannot be confirmed.

Virtual Classroom Guidelines

Please note the following standards about our course's online learning community:
1. (**Academic Conduct**) Remember that your participation in this course, via virtual discussions in written or video form, must abide by the university's policies and procedures. Students engaging in disrespectful behavior will be removed from the course.
2. (**Use of AI**) Online tools (such as ChatGPT) can be hugely helpful in this course. Students are encouraged to leverage such tools for checking their answers, improving segments of code (such as helper functions or graphing utilities), or creating study guides. Submitting entire homework solutions or sections of your final report that are created entirely by AI tools is prohibited and is considered academic misconduct.
3. (**Professor Created Videos**) The web-based video delivery of each class in this course is for sole use of the students enrolled in this course. Any other use of these class videos or any pictures or derivatives of the class videos without the written consent of the course's professor is prohibited.
4. (**Student Created Videos**) The videos created by students as part of this course are for sole use of the students enrolled in this course. Any other use of these videos or any pictures or derivatives of the videos without the written consent of the video creator is prohibited.

Copyright

My lectures, notes, handouts, and displays are protected by state common law and federal copyright law. They are my own original expression, based in part on the sources identified in the course notes, and I've recorded them prior or during my lecture in order to ensure that I obtain copyright protection. Students are authorized to take notes in my class; however, this authorization extends only to making one set of notes for your own personal use and no other use. I will inform you as to whether you are authorized to record my lectures at the beginning of each semester. If you are so authorized to record my lectures, you may not copy this recording or any other material, provide copies of either to anyone else, or make a commercial use of them without prior permission from me.

Student Responsibilities and Resources

As a member of the University of Connecticut student community, you are held to certain standards and academic policies. In addition, there are numerous resources available to help you succeed in your academic work.

Note: The use of AI as a tool in this course is encouraged, particularly to assist in coding assignments. As the student, you are responsible for ensuring that (1) submissions to any assignments are in your own words, and (2) you understand the underlying concepts in any situation where AI is used for assistance.

Review these important standards, policies and resources, which include:

- The Student Code
  - Resources on Avoiding Cheating and Plagiarism
- Academic, Scholarly, and Professional Integrity and Misconduct (ASPIM)
- Copyrighted Materials
- Credit Hours and Workload
- Netiquette and Communication
- Adding or Dropping a Course
- Academic Calendar
- Policy Against Discrimination, Harassment and Inappropriate Romantic Relationships
- Sexual Assault Reporting Policy

Student Health and Wellness

The University of Connecticut strives to support the optimal well-being of all students. Student Health and Wellness (SHaW) offers a comprehensive set of services including medical care, mental health, and health promotion.

The University of Connecticut is committed to protecting the rights of individuals with disabilities and assuring that the learning environment is accessible. Students who require accommodations should contact the Center for Students with Disabilities, Wilbur Cross Building Room 204, (860) 486-2020 or http://csd.uconn.edu/.

Blackboard measures and evaluates accessibility using two sets of standards: the WCAG 2.0 standards issued by the World Wide Web Consortium (W3C) and Section 508 of the Rehabilitation Act issued in the United States federal government." (Retrieved March 24, 2013 from Blackboard's website)

Software/Technical Requirements (with Accessibility and Privacy Information)

The University has set minimum device requirements for all students. **NOTE:** Chromebooks do not meet the minimum requirements.

The software/technical requirements for this course include:

- HuskyCT/Blackboard (HuskyCT/ Blackboard Accessibility Statement, HuskyCT/ Blackboard Privacy Policy)
- Adobe Acrobat Reader (Adobe Reader Accessibility Statement, Adobe Reader Privacy Policy)
- Microsoft Office (free to UConn students) (Microsoft Accessibility Statement, Microsoft Privacy Statement)
- Dedicated access to high-speed internet with a minimum speed of 1.5 Mbps (4 Mbps or higher is recommended).
- Webcam
- Python with Anaconda (see https://intro-stat-learning.github.io/ISLP/installation.html)

For information on managing your privacy at the University of Connecticut, visit the University's Privacy page.

NOTE: This course has NOT been designed for use with mobile devices.

Help

This course is facilitated online using the learning management platform, HuskyCT. The IT Knowledge Base provides students with support, troubleshooting, and how-to information about HuskyCT. The IT Knowledge Base includes a video tour of HuskyCT.

For technical help with HuskyCT, you have access to the in-person/live person support options available during regular business hours through the Technology Support Center. You also have 24x7 Course Support outside of business hours, including access to live chat, phone, and support documents.

Technical and Academic Help provides a guide to frequently asked questions for online students.

Study Groups

Are you interested in forming a study group with other students in the class? There is a study group application in Nexus that can help you get started. View this video for more information.

Minimum Technical Skills

To be successful in this course, you will need the following technical skills:
- Use electronic mail with attachments.
- Save files in commonly used word processing program formats.
- Copy and paste text, graphics or hyperlinks.
- Work within two or more browser windows simultaneously.
- Access and create PDF files.
- Use a webcam and microphone.

Evaluation of the Course

Students will be provided an opportunity to evaluate instruction in this course using the University's Student Evaluation of Teaching (SET), which is administered by the Office of Budget, Planning and Institutional Research (BPIR).

Additional informal formative surveys may also be administered within the course as an optional evaluation tool.