# C S 329E : Elements of Data Analytics (52740 & 52745)

Fall 2021

---

**Class Meets:** MWF 8am-9am GDC 1.304 or synchronously via zoom (52740)
MWF 10am-11am UTC 3.124 or synchronously via zoom (52745)

**Course Mode:** *Hybrid (utilizes both online and in-person experiences)*


**Instructor**: Anna Chaney
**Pronouns:** She/her
**Email:** chaney@utexas.edu

**Office:** GDC 4.306
**Office Hours**: 11:00am – 12:15pm Wednesday

**Teaching Assistant (**52740**)**: Yan Zheng
**Pronouns**: she/her
**Email:** yanzheng@utexas.edu

**Office Hours**: Thursdays 5-6pm Basement of GDC, room 1.302, Desk 1

**Teaching Assistant (**52745**)**: Taijing Chen
**Pronouns**: she/her
**Email**: taijingchen@utexas.edu

**Office Hours**: Office hours Friday 1-2pm Basement of GDC


## Course Description

### Basic Catalog Information

An exploration of algorithms to find patterns in datasets. Topics include regression, classification, clustering, anomaly detection, and association analysis. Three lecture hours a week for one semester. May not be counted toward a degree in computer science. May be repeated for credit when the topics vary. Prerequisite: Computer Science 313E, 314, or 314H with a grade of at least C-.

**Program**: Undergraduate Program
**Division**: Non-Majors

### Course Objective

Data focused problem solving has emerged as a preeminent technique to infuse computer systems with skills traditionally left to human brains. Making predictions about what things are, or what they may become is something our society does continuously – from recognizing our family member's voice on the telephone to approving a loan application. Computationally programming computers to help us make this style of prediction has theoretical underpinnings that are decades old, and more recently have become practical solutions to problems through the advancement in computer hardware.

We'll cover a 365° view of how computer scientists (or, engineers in general) choose and implement data focused algorithms, how to conduct experiments to compare the results of different algorithms, visualizing results, how to write code in a common data science computing platform, and what implications that data and algorithm selection has on society.

### Learning Goals

*Practice* cleaning and manipulating data
*Gain* experience writing python code using common packages
*Develop* an understanding of data mining algorithms
*Evaluate* the performance of data mining algorithms
*Collaborate* with your colleagues
*Navigate* and *understand* technical documentation

### How to succeed in this course

1. **Practice**. Practice is essential to developing the skills you are learning in this class. It also actually helps you learn, because some things which seem murky clarify when you actually do them, and sometimes trying to do something shows that you only thought you understood it.
2. **Participate**. By seeing what you can and cannot do, and what comes easily and what you struggle with, I can help you learn better, by giving advice and, if need be, adjusting the course. Ask questions in discussion forums, and if you see a peer asking a question that you might be able to help with, participate by helping your fellow student.
3. **Present**. The university is, in the end, going to stake its reputation (and that of its faculty) on assuring the world that you have mastered the skills and learned the material that goes with your degree. With this in mind, treat each homework assignment and project like a presentation that will be judged by its cohesiveness, and ability to stand alone as a piece of data analysis.

## Course Requirements

### Required Textbooks

- *Introduction to Data Mining (2nd Edition)* by Tan, Steinbach, Karpatne, and Kumar
    - *"Introduction to Data Mining" for this class is available through the Longhorn Textbook Access (LTA) program, a new initiative between UT Austin, The University Co-op and textbook publishers to significantly reduce the cost of digital course materials for students. You can access your required materials through the "My Textbooks" tab in Canvas. You are automatically opted into the program but can easily opt-out (and back in) via Canvas through the 12th class day. If you remain opted-in at the end of the 12th class day you will receive a bill through your "What I Owe" page and have until the end of the 18th class day to pay and retain access. If you do not pay by the 18th class day, you will lose access to the materials after the 20th class day and your charge will be removed. More information about the LTA program is available at universitycoop.com/longhorn-textbook-access.*
- *The Elements of Statistical Learning* by Hastie, Tibshirani, and Friedman. Available for free download via https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf

### Recommended books (supplemental material which is not required)

- *Weapons of Math Destruction* by Cathy O'Neil.
- *Pattern Recognition and Machine Learning* by Bishop. Available for free download via https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf
- *Machine Learning, A Probabilistic Approach,* by Murphy

### Required Devices

It is assumed that all students will have access to a computer to participate in Zoom class meetings, access course materials on Canvas, turn in assignments on Canvas, and take exams on GradeScope. Assignments will require creating virtual python environments with Anaconda, installing python packages, and creating visualizations as necessary for the course as described in the assignments. If you do not have access to a personal computer, please communicate with me as soon as possible so we may discuss how the computer science department can support you.

## Classroom expectations

Professional conduct is built upon the idea of mutual respect. Such conduct entails (but is not necessarily limited to):

**Attending class**: The class benefits from the attendance and participation of all students. This class is listed as hybrid, which means we can meet in the classroom or on-line via zoom. Class attendance is mandatory, and I will give you at least one week notice of when we are meeting on-line vs in person in the classroom. When meeting via zoom please plan to have your video on during class and be certain that your display name is your actual name (the one you use in person).

**Arriving on time**: Please do not hesitate to come to class, even if you are arriving late. I would rather you attend some of class, than miss out altogether. However, if you must arrive late or leave early, please be considerate of others.

**Minimizing disruptions**: You are invited to participate in class synchronously regardless of the noise-level of your location or the number of people in your background (or foreground). However, please do keep yourself on mute during lecture so as to make sure all can hear. Note that the spacebar can be used to temporarily unmute yourself during Zoom meetings (press to unmute, release to re-mute) on most operating systems.

During drop-in hours and discussion section, please keep mute turned off to allow for a more natural conversation---unless noise levels are high in your location.

**Respect**: You should act respectfully toward all class participants.

**Policy on Children in Class/on Zoom** Children are always welcome. I understand that if you have childcare responsibilities you may be stretched thin. Please talk to me if you need help.

## Grading for this Course

| Category | Percentage of Grade | Notes |
|---|---|---|
| Class Participation Quizzes | 15% | Synchronous |
| Homework Assignments | 55% | Lowest HW score will be dropped |
| Exam 1 | 10% | 29 Sept 2021 |
| Exam 2 | 10% | 27 Oct 2021 |
| Exam 3 | 10% | 6 Dec 2021 |

## Course Outline

All instructions, assignments, readings, rubrics and essential information will be on the Canvas website at https://utexas.instructure.com. Check this site regularly and use it to ask questions about the course schedule.

**Changes** to the schedule may be made at my discretion and if circumstances require. It is your responsibility to note these changes when announced (although I will do my best to ensure that you receive the changes with as much advanced notice as possible).

| Date | Week | Topic Family | Lecture | What is Due |
|---|---|---|---|---|
| 25-Aug-21 | 1 | Intro | Introduction to the course | |
| 27-Aug-21 | | Python for Data Science | Jupyter Notebooks and Packages | |
| 30-Aug-21 | 2 | | Pandas / Matplotlib | |
| 1-Sep-21 | | | Design of Experiments | |
| 3-Sep-21 | | Math Prereq | Probability and Stats and Set Theory and Linear Algebra | HW #1 |
| 6-Sep-21 | 3 | | LABOR DAY - NO CLASSES | |
| 8-Sep-21 | | | Probability and Stats and Set Theory and Linear Algebra | |
| 10-Sep-21 | | Regression | Regression | |
| 13-Sep-21 | 4 | | Regression | HW #2 |
| 15-Sep-21 | | | Feature Engineering | |
| 17-Sep-21 | | | Feature Engineering | |
| 20-Sep-21 | 5 | Classification | Decision Trees | HW #3 |
| 22-Sep-21 | | | Decision Trees | |
| 24-Sep-21 | | | Overfitting and Cross Validation | |
| 27-Sep-21 | 6 | | Overfitting and Cross Validation | HW #4 |
| 29-Sep-21 | | | Exam 1 | |
| 1-Oct-21 | | | Nearest Neighbor | |
| 4-Oct-21 | 7 | | Nearest Neighbor | HW #5 |
| 6-Oct-21 | | | Naïve Bayes | |
| 8-Oct-21 | | | Naïve Bayes | |
| 11-Oct-21 | 8 | | Evaluation of Classifiers | HW #6 |
| 13-Oct-21 | | | Evaluation of Classifiers | |
| 15-Oct-21 | | | Ensemble Methods | |
| 18-Oct-21 | 9 | | Ensemble Methods | HW #7 |
| 20-Oct-21 | | | SVMs | |
| 22-Oct-21 | | | Neural Networks | |
| 25-Oct-21 | 10 | | Neural Networks | HW #8 |
| 27-Oct-21 | | | Exam 2 | |
| 29-Oct-21 | | Clustering | K-Means | |
| 1-Nov-21 | 11 | | K-Means | |
| 3-Nov-21 | | | Density Based Clustering | |
| 5-Nov-21 | | | DB activity | HW #9 |
| 8-Nov-21 | 12 | | Hierarchical Clustering | |
| 10-Nov-21 | | | Hier. Activity | |
| 12-Nov-21 | | | Evaluating Clusters | HW #10 |
| 15-Nov-21 | 13 | Anomaly Detection | Anomaly Detection | |
| 17-Nov-21 | | | Anomaly Detection | |
| 19-Nov-21 | | Association Analysis | Association Analysis - Apriori | HW #11 |
| 22-Nov-21 | 14 | | Rule Gen | |
| 24-Nov-21 | | | THANKSGIVING BREAK - NO CLASSES | |
| 26-Nov-21 | | | THANKSGIVING BREAK - NO CLASSES | |
| 29-Nov-21 | 15 | | Compact Itemsets / Skewed Distributions | |
| 1-Dec-21 | | | Evaluating Association Analysis | |
| 3-Dec-21 | | | Review | HW #12 |
| 6-Dec-21 | 16 | | Exam 3 | |

# Policies

## Classroom Policies

**Statement on Learning Success**

Your success in this class is important to me. We will all need accommodations because we all learn differently. If there are aspects of this course that prevent you from learning or exclude you, please let me know as soon as possible. Together we'll develop strategies to meet both your needs and the requirements of the course. I also encourage you to reach out to the student resources available through UT. Many are listed on this syllabus, but I am happy to connect you with a person or Center if you would like.

**Grading Policies**

I have instituted some policies that will help you if a particular day, or assignment has gone particularly poorly for you. One homework assignment will be dropped completely, and up to two homework assignments can be late. I reserve the right to assign grades on a curve, but at the very minimum I will promise these grade cut-offs:

| Grade | Cutoff |
|---|---|
| A<br>A- | 94.0%<br>90.0% |
| B+<br>B<br>B- | 87.0%<br>84.0%<br>80.0% |
| C+<br>C<br>C- | 77.0%<br>74.0%<br>70.0% |
| D | 65.0% |
| F | <65.0% |

**GRACE POLICY: Homework Time-bank options**

Sometimes we have bad days, bad weeks, and bad semesters. In an effort to accommodate any unexpected, unfortunate personal crisis, I have built "grace days" into our course. You do not have to utilize this policy, but if you find yourself struggling with unexpected personal events, I encourage you to e-mail me as soon as possible to notify me that you are using our grace policy. You may use this policy one of two ways (please choose, and let me know):
You may have a two-day grace period for one homework assignment, OR
You may have 2 one-day extensions for two different homework assignments.
Homework will not be accepted more than two days late without speaking to Prof Chaney regarding the extenuating circumstances.
If you have exhausted your grace days, you may still turn in the homework up to two days late with a 10% deduction per late day.

**Regrading Policy**

It is very important to us that all assignments are properly graded. If you believe there is an error in your assignment grading, please submit an explanation via email to Prof Chaney and the TA within 7 days of receiving the grade. No regrade requests will be accepted orally, and no regrade requests will be accepted more than 7 days after you receive the grade for the assignment.

**Absences**

You are responsible for all participation activities assigned during the class, even if you can not attend the class synchronously. Please check Canvas.

## Student Rights & Responsibilities

- You have a right to a learning environment that supports mental and physical wellness.
- You have a right to respect.
- You have a right to be assessed and graded fairly.
- You have a right to freedom of opinion and expression.
- You have a right to privacy and confidentiality.
- You have a right to meaningful and equal participation, to self-organize groups to improve your learning environment.
- You have a right to learn in an environment that is welcoming to all people. No student shall be isolated, excluded or diminished in any way.

*With these rights come responsibilities:*

- You are responsible for taking care of yourself, managing your time, and communicating with the teaching team and with others if things start to feel out of control or overwhelming.
- You are responsible for acting in a way that is worthy of respect and always respectful of others.
- Your experience with this course is directly related to the quality of the energy that you bring to it, and your energy shapes the quality of your peers' experiences.
- You are responsible for creating an inclusive environment and for speaking up when someone is excluded.
- You are responsible for holding yourself accountable to these standards, holding each other to these standards, and holding the teaching team accountable as well.

**Personal Pronoun Use (She / He / They / Ze / Etc)**

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. Class rosters are provided to the instructor with the student's legal name, unless they have added a "preferred name" with the Gender and Sexuality Center (http://diversity.utexas.edu/genderandsexuality/publications-and-resources/). I will gladly honor your request to address you by a name that is different from what appears on the official roster, and by the gender pronouns you use (she/he/they/ze, etc). Please advise me of any changes early in the semester so that I may make appropriate updates to my records.

**University Resources for Students**

*"Keep Learning" Resources*

This course may be offered in a format to which you are unaccustomed. If you are looking for ideas and strategies to help you feel more comfortable participating in our class, please explore the resources available here: https://onestop.utexas.edu/keep-learning/

*Services for Students with Disabilities*

This class respects and welcomes students of all backgrounds, identities, and abilities. If there are circumstances that make our learning environment and activities difficult, if you have medical information that you need to share with me, or if you need specific arrangements in case the building needs to be evacuated, please let me know. I am committed to creating an effective learning environment for all students, but I can only do so if you discuss your needs with me as early as possible. I promise to maintain the confidentiality of these discussions. Any student with a documented disability who requires academic accommodations should contact Services for Students with Disabilities at 471-6259 (voice) or 512-410-6644 (Video Phone) as soon as possible to request an official letter outlining authorized accommodations. For more information, visit http://ddce.utexas.edu/disability/about/.

*Counseling and Mental Health Center*

Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. http://www.cmhc.utexas.edu/individualcounseling.html

*The Sanger Learning Center*

Did you know that more than one-third of UT undergraduate students use the Sanger Learning Center each year to improve their academic performance? All students are welcome to take advantage of Sanger Center's classes and workshops, private learning specialist appointments, peer academic coaching, and tutoring for more than 70 courses in 15 different subject areas. For more information, please visit http://www.utexas.edu/ugs/slc or call 512-471-3614 (JES A332).

*Undergraduate Writing Center*: http://uwc.utexas.edu/
*Libraries*: http://www.lib.utexas.edu/
*ITS*: http://www.utexas.edu/its/
*Student Emergency Services*: http://deanofstudents.utexas.edu/emergency/

*BeVocal*

BeVocal is a university-wide initiative to promote the idea that individual Longhorns have the power to prevent high-risk behavior and harm. At UT Austin all Longhorns have the power to intervene and reduce harm. To learn more about BeVocal and how you can help to build a culture of care on campus, go to: https://wellnessnetwork.utexas.edu/BeVocal.

**Important Safety Information:**

If you have concerns about the safety or behavior of fellow students, TAs or Professors, call BCAL (the Behavior Concerns Advice Line): 512-232-5050. Your call can be anonymous. If something doesn't feel right – it probably isn't. Trust your instincts and share your concerns.

The following recommendations regarding emergency evacuation from the Office of Campus Safety and Security, 512-471-5767, http://www.utexas.edu/safety/

Occupants of buildings on The University of Texas at Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside.

- Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building.
- Students requiring assistance in evacuation shall inform their instructor in writing during the first week of class.
- In the event of an evacuation, follow the instruction of faculty or class instructors. Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office.
- Link to information regarding emergency evacuation routes and emergency procedures can be found at: www.utexas.edu/emergency

**Title IX Reporting**

Title IX is a federal law that protects against sex and gender-based discrimination, sexual harassment, sexual assault, sexual misconduct, dating/domestic violence and stalking at federally funded educational institutions. UT Austin is committed to fostering a learning and working environment free from discrimination in all its forms. When sexual

misconduct occurs in our community, the university can:

1. Intervene to prevent harmful behavior from continuing or escalating.
2. Provide support and remedies to students and employees who have experienced harm or have become involved in a Title IX investigation.
3. Investigate and discipline violations of the university's relevant policies (https://titleix.utexas.edu/relevant-polices/).

Beginning January 1, 2020, Texas Senate Bill 212 requires all employees of Texas universities, including faculty, report any information to the Title IX Office regarding sexual harassment, sexual assault, dating violence and stalking that is disclosed to them. Texas law requires that all employees who witness or receive any information of this type (including, but not limited to, writing assignments, class discussions, or one-on-one conversations) must be reported. **I am a Responsible Employee and must report any Title IX related incidents** that are disclosed in writing, discussion, or one-on-one. Before talking with me, or with any faculty or staff member about a Title IX related incident, be sure to ask whether they are a responsible employee. If you would like to speak with someone who can provide support or remedies without making an official report to the university, please email advocate@austin.utexas.edu. For more information about reporting options and resources, visit http://www.titleix.utexas.edu/, contact the Title IX Office via email at titleix@austin.utexas.edu, or call 512-471-0419.

Although graduate teaching and research assistants are not subject to Texas Senate Bill 212, they are still mandatory reporters under Federal Title IX laws and are required to report a wide range of behaviors we refer to as sexual misconduct, including the types of sexual misconduct covered under Texas Senate Bill 212. The Title IX office has developed supportive ways to respond to a survivor and compiled campus resources to support survivors.

**Emergency Evacuation Procedures**

The following recommendations regarding emergency evacuation from the Office of Campus Safety and Security, 512-471-5767, http://www.utexas.edu/safety/

## University Policies

**Academic Integrity**

Each student in the course is expected to abide by the University of Texas Honor Code: "As a student of The University of Texas at Austin, I shall abide by the core values of the University and uphold academic integrity." **Plagiarism is taken very seriously at UT.** Therefore, if you use words or ideas that are not your own (or that you have used in previous class), you must cite your sources. Otherwise you will be guilty of plagiarism and subject to academic disciplinary action, including failure of the course. You are responsible for understanding UT's Academic Honesty and the University Honor Code which can be found at the following web address: https://deanofstudents.utexas.edu/conduct/standardsofconduct.php

**Q Drop Policy**

If you want to drop a class after the 12th class day, you'll need to execute a Q drop before the Q-drop deadline, which typically occurs near the middle of the semester. Under Texas law, you are only allowed six Q drops while you are in college at any public Texas institution. For more information, see: http://www.utexas.edu/ugs/csacc/academic/adddrop/qdrop

**COVID-19 Updates: Fall 2021 Semester**

Classroom Safety and COVID-19

To help preserve our in person learning environment, the university recommends the following.

- Adhere to university mask guidance.
- Vaccinations are widely available, free and not billed to health insurance. The vaccine will help protect against the transmission of the virus to others and reduce serious symptoms in those who are vaccinated.
- Proactive Community Testing remains an important part of the university's efforts to protect our community. Tests are fast and free.

Visit utexas.edu for more information.

Sharing of Course Materials is Prohibited:

No materials used in this class, including, but not limited to, lecture hand-outs, videos, assessments (quizzes, exams, papers, projects, homework assignments), in-class materials, review sheets, and additional problem sets, may be shared online or with anyone outside of the class unless you have my explicit, written permission. Unauthorized sharing of materials promotes cheating. It is a violation of the University's Student Honor Code and an act of academic dishonesty. I am well aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to Student Conduct and Academic Integrity in the Office of the Dean of Students. These reports can result in sanctions, including failure in the course.

Class Recordings:

Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.