

# C S 329E : Elements of Data Analytics (50985 & 50989)

Fall 2020

---

**Class Meets:** MWF 9am-10am (50985) / MWF 11am-12pm (50989) Central Time on Zoom

**Course Mode:** *Internet*

**Instructor:** Anna Chaney

**Pronouns:** She/her

**Email:** [chaney@cs.utexas.edu](mailto:chaney@cs.utexas.edu)

**Teaching Assistants:** Yan Zheng

**Pronouns:** She/her

**Email:** [yanzheng@utexas.edu](mailto:yanzheng@utexas.edu)

**Office:** GDC 4.306 (Office hours on Zoom)

**Office Hours:** times on Canvas

**Office:** Zoom

**Office Hours:** times on Canvas

## Course Description

### Basic Catalog Information

An exploration of machine learning algorithms to find patterns in large datasets. Topics include 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 the 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.

### 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 (2<sup>nd</sup> Edition)* by Tan, Steinbach, Karpapne, and Kumar
- *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](https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf)

### Recommended books

- *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, so 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). However, we understand that there may be extenuating circumstances that prevent you from attending synchronously. Therefore, we will provide for both asynchronous and synchronous participation options. There will be a class participation grade.

**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.

In this environment, some disruptions are unavoidable---my own family will likely disrupt early and often---and that's okay. Please do not allow fear of disruptions to prevent you from attending.

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

**Appearance:** Please arrive at class events clothed. You don't have to look great, and we can't smell you, but being covered is not optional. (Obvious nightwear is also not appropriate.)

**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	10%	Will have asynchronous and synchronous options
Homework Assignments	55%	Lowest HW score will be dropped
Exam 1	10%	
Exam 2	10%	
Final Exam	15%	Cumulative

## 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	Activity	DM Textbook	What is Due
26-Aug	1	Introduction	Introduction to the course		
28-Aug		Python Basics	Introduction to python		
31-Aug	2		Pandas / Matplotlib		
2-Sep		Math Basics	Probability and Stats		HW #1
4-Sep			Probability and Stats		
7-Sep	3		LABOR DAY - NO CLASS		
9-Sep		Classification	Regression		HW #2
11-Sep			Regression		
14-Sep	4		Regression		
16-Sep			Decision Trees	Ch 3.1 - 3.3	HW #3
18-Sep			Decision Trees		
21-Sep	5		Decision Trees		
23-Sep			Overfitting and Cross Validation	Ch 3.4 - 3.9	HW #4
25-Sep			Overfitting and Cross Validation		
28-Sep	6		Exam		Exam 1
30-Sep			Design of Experiments		
2-Oct			Nearest Neighbor	Ch 4.3	
5-Oct	7		Nearest Neighbor		
7-Oct			Naïve Bayes	Ch 4.4	HW #5
9-Oct			Naïve Bayes		
12-Oct	8		Evaluation of Classifiers	Ch 4.11	
14-Oct			Ensemble Methods	Ch 4.10	HW #6
16-Oct			SVMs	Ch 4.9	
19-Oct	9		SVMs		
21-Oct			Neural Networks	Ch 4.7	HW #7
23-Oct			Neural Networks		
26-Oct	10		Review		
28-Oct			Exam		Exam 2
30-Oct		Clustering	K-Means	Ch 7.1 - 7.2	
2-Nov	11		K-Means		
4-Nov			Density Based Clustering	Ch 7.4	HW #8
6-Nov			Hierarchical Clustering	Ch 7.3	
9-Nov	12		Evaluating Clusters	Ch 7.5	
11-Nov			Anomaly Detection	Ch 9	HW #9
13-Nov		Association Analysis	Apriori	Ch 5.1-5.5	
16-Nov	13		Scalability Issues & Rule Generation	Ch 5.3	
18-Nov			FP Growth	Ch 5.6	HW #10
20-Nov			Compact Itemsets / Skewed Distributions	Ch 5.4	
23-Nov	14		Evaluating Association Patterns	Ch 5.7-5.8	
25-Nov			Sequential Patterns	Ch 6.4	HW #11
27-Nov			NO CLASS - Thanksgiving		
30-Nov	15		NO CLASS - Thanksgiving		
2-Dec			Exam Review		
4-Dec			Exam		Final

## 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	94%
A-	90%
B+	87%
B	84%
B-	80%
C+	77%
C	74%
C-	70%
D	65%
F	<65%

#### GRACE POLICY: 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 "time banks" 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 assignment, OR  
You may have 2 one-day extensions for two different assignments.

#### 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**

#### **COVID-19 Update: "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:**

**COVID-19 Update:** While we will post information related to the contemporary situation on campus, you are encouraged to stay up-to-date on the latest news as related to the student experience.

<https://coronavirus.utexas.edu/students>

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](http://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-policies/) (<https://titleix.utexas.edu/relevant-policies/>).

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](mailto: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](mailto: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/standardssofconduct.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 2020 Semester

### Safety and Class Participation/Masks:

For every face-to-face class experience, we will all need to make some adjustments in order to benefit from in-person classroom interactions in a safe and healthy manner. Our best protections against spreading COVID-19 on campus are masks (defined as cloth face coverings) and staying home if you are showing symptoms. Therefore, for the benefit of everyone, this means that all students are required to follow two important rules.

- **Every student must wear a cloth face covering properly in class and in all campus buildings at all times.**
- **Every student must engage in documented daily symptom screening.** This means that each class day in which on campus activities occur, students must upload certification from the symptom tracking app and confirm that they completed their symptom screening for that day to Canvas. Students should not upload the results of that screening, just the certificate that they completed it. If the symptom tracking app recommends that the student isolate rather than coming to class, then students must not return to class until cleared by a medical professional.

If a student is not wearing a cloth face covering properly in the classroom (or any UT building), that student must leave the classroom (and building). If the student refuses to wear a cloth face covering, class will be dismissed for the remainder of the period, and the student will be subject to disciplinary action as set forth in the university's Institutional Rules/General Conduct 11-404(a)(3). Students who have a condition that precludes the wearing of a cloth face covering must follow the procedures for [obtaining an accommodation](https://orientation.utexas.edu/students-with-disabilities) (<https://orientation.utexas.edu/students-with-disabilities>)

### 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.