

# INTRODUCTION TO DATA MINING

CS 363D, SPRING 2018, 51593  
TU/TH 12:30-2:00, PAR 203



## INSTRUCTOR

Angie Beasley  
angie.beasley@utexas.edu  
Office Hours: Tues 11:00-12:00  
Thurs 2:00-3:00  
GDC 4.314

## TA

Pratyush Kar  
pkar@cs.utexas.edu  
Office Hours: Mon 3:30-5:30pm  
GDC 1.302



## PROCTOR

Nandhini Lakuduva  
nlakuduva@utexas.edu  
Office Hours: Wed 9:30-11:30am  
GDC 1.302

## COURSE DESCRIPTION



This course is an introduction to the topic of data mining. We will cover data preprocessing, classification, clustering, anomaly detection, and association analysis.

---

Computer Science 363D and 378 (Topic: Introduction to Data Mining) may not both be counted.

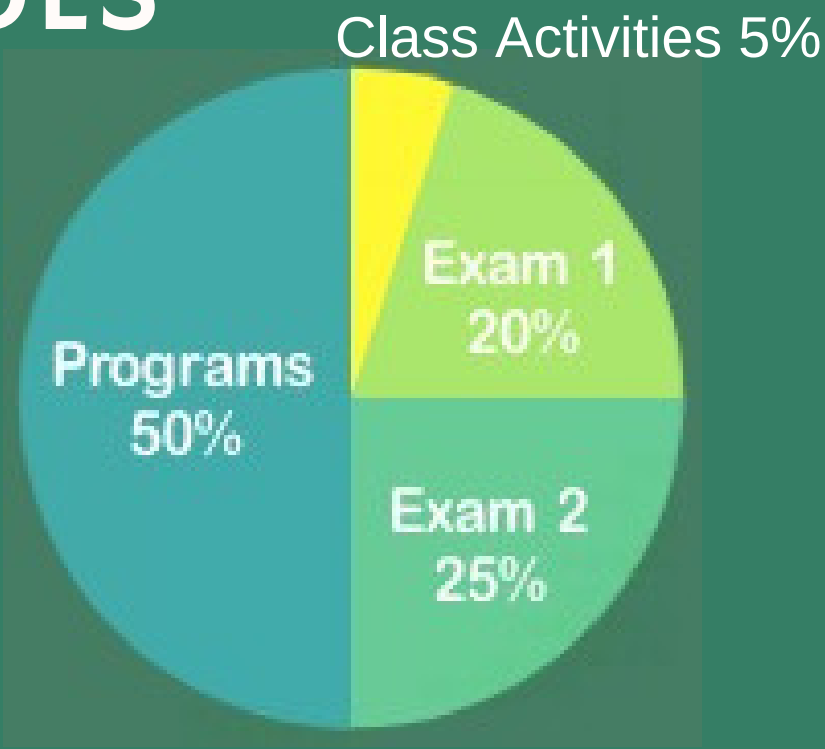
Prerequisites: The following coursework with a grade of at least C- : Computer Science 429 (or 310) or 429H (or 310H); Mathematics 362K or Statistics and Data Sciences 321 (or Statistics and Scientific Computation 321); and Mathematics 340L, 341, or Statistics and Data Sciences 329C (or Statistics and Scientific Computation 329C).

## TEXTBOOK

Introduction to Data Mining  
by Pang-Ning Tan  
Michael Steinbach  
Vipin Kumar



# GRADES



A	>94
A-	90-93
B+	87-89
B	84-86
B-	80-83
C+	77-79
C	74-76
C-	70-73
D+	67-69
D	64-66
D-	60-63
F	<59

All numbers are absolute and will not be rounded up or down at any stage.



## IN-CLASS ACTIVITIES

Throughout the semester, there will be in-class activities. They will vary in nature and will be random in occurrence. Some will be graded for correctness, and some will only be graded for completion. You may drop your 2 lowest of these and the remaining will make up 5% of your final grade.

## PROGRAMMING ASSIGNMENTS

There will be 8-10 programming assignments, each equally weighted to total 50% of your final grade.



Programming assignments will be completed using Python 3 and Jupyter Notebooks. Assignments may be worked individually or in pairs. If you work in pairs, you are expected to use the proper pair programming method.

## LATE ASSIGNMENTS

You will have 4 late days in 1-day units (that is, 1 minute to 24 hours late = 1 late day) to use throughout the semester. You may divide your late days across the programming assignments in any way you wish. Once you have used all of your late days, late assignments will no longer be accepted.

In the case of pair programming, each member of the pair must have enough late days to cover the late submission. So if the pair submits their code 2 days late, each member must have two late days remaining to use and each member will lose two late days.

To use late days, you only need to submit the assignment. You do not need to email the instructor or the TA, you do not need to indicate that you are using late days. Your late days will be deducted according to when your assignment is submitted. If you submit a late assignment without enough late days to support it, you will receive a zero for that assignment.

Contact me if there are extenuating circumstances.

# COURSE SCHEDULE

*Subject to change at instructor's discretion.*

- 1/16 (Ice Day - No class)
- 1/18 Introduction
- 1/23 Data Preprocessing [Ch 1-3]

## CLASSIFICATION

- 1/25 Decision Trees [Ch 4.1 - 4.3]
- 1/30 Decision Trees (cont.)
- 2/1 Overfitting [Ch 4.4]
- 2/6 Nearest Neighbor [Ch 5.2]
- 2/8 Naive Bayes [Ch 5.3]
- 2/13 Evaluating Classifiers [Ch 4.5-4.6, 5.7-5.8]
- 2/15 Ensemble Methods [Ch 5.6]
- 2/20 TBD - Career Fair
- 2/22 SVMs [Ch 5.5]
- 2/27 Neural Nets [Ch 5.4]
- 3/1 Neural Nets (cont.)
- 3/6 Exam Review
- 3/8 **EXAM 1**

## CLUSTERING

- 3/13 Spring Break
- 3/15 Spring Break
- 3/20 Clustering & K-means [Ch 8.1 - 8.2]
- 3/22 Density-Based & Fuzzy Clustering [Ch 8.4, 9.2]
- 3/27 Hierarchical Clustering [Ch 8.3]
- 3/29 Evaluating Clusters [Ch 8.5]
- 4/1 Anomaly Detection [Ch 10]

## ASSOCIATION ANALYSIS

- 4/5 Apriori [Ch 6.1-6.5]
- 4/10 Scalability Issues
- 4/12 FP Growth [Ch 6.6]
- 4/17 Compact Itemsets/Skewed Distributions [6.4, 6.9]
- 4/19 Evaluating Association Patterns [6.7-6.8]
- 4/24 Sequential Patterns [Ch 7.4]
- 4/26 Infrequent Patterns [Ch 7.6]
- 5/1 Guest Speaker (tentative)
- 5/4 Exam Review
- 5/11 **EXAM 2**

# 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.”

This means that work you produce on assignments and exams is all your own work, unless it is assigned as group work. I will make it clear for each exam or assignment whether collaboration is allowed or not.

You are responsible for understanding UT’s Academic Honesty Policy which can be found here:

[http://deanofstudents.utexas.edu/sjs/acint\\_student.php](http://deanofstudents.utexas.edu/sjs/acint_student.php)



**If you submit code that is not your own, you will be guilty of plagiarism and subject to academic disciplinary action, including failure of the course.**

## ANONYMOUS FEEDBACK

Anonymous feedback may be provided to the instructor at anytime via

Canvas -> Quizzes -> Anonymous Feedback

## UNIVERSITY RESOURCES

The Counseling and Mental Health Center (CMHC) provides counseling, psychiatric, consultation, and prevention services: <http://cmhc.utexas.edu/>

Student Emergency Services

<http://deanofstudents.utexas.edu/emergency/>

Need help with technology? <http://www.utexas.edu/its/>

Canvas help is available 24/7 at

<https://utexas.instructure.com/courses/633028/pages/student-tutorials>

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.



# UNIVERSITY POLICIES

## RELIGIOUS HOLY DAYS

By UT Austin policy, you must notify me of your pending absence at least fourteen days prior to the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, I will give you an opportunity to complete the missed work within a reasonable time after the absence.

## 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>

## STUDENT ACCOMMODATIONS

Students with a documented disability may request appropriate academic accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 512-471-6259 (voice) or 1-866-329-3986 (video phone).

<http://ddce.utexas.edu/disability/about/>

Please request a meeting with me as soon as possible to discuss any accommodations you may need.

Please notify me as soon as possible if the material being presented in class is not accessible to you.

Please notify me as soon as possible if any of the physical space is difficult for you.

## EVACUATION INFORMATION

The Office of Campus Safety and Security:

<http://www.utexas.edu/safety/>

Occupants of buildings on The University of Texas at Austin campus are required to evacuate buildings when an alarm or alert is activated. Exit and assemble outside, unless told otherwise by an official representative. Do not re-enter a building unless given instructions by the Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office.

- 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 for entry.
- Students requiring assistance in evacuation shall inform their instructor in writing during the first week of class.
- Information regarding emergency evacuation routes and emergency procedures can be found at:  
[www.utexas.edu/emergency](http://www.utexas.edu/emergency)