

INTRODUCTION TO DATA MINING

CS 363D, FALL 2017, 51709
TU/TH 12:30-2:00, GDC 5.302



INSTRUCTOR


Angie Beasley
angie.beasley@utexas.edu
Office Hours: Tues 11:00-12:30
Thurs 2:00-3:00
GDC 4.314

TA

Pratyush Kar
pkar@cs.utexas.edu
Office Hours: Wed 5:00-7:00pm
GDC 1.302, Desk 1



COURSE DESCRIPTION



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

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



Programming : 50% Exams: 50%
Program 0: 5% Exam 1: 16.6%
Program 1: 15% Exam 2: 16.6%
Program 2: 15% Exam 3: 16.6%
Program 3: 15%

Final grades will be determined on the basis of this rubric. Note: all numbers are absolute and will not be rounded up or down at any stage.

< 59	60-63	64-66	67-69	70-73	74-76	77-79	80-83	84-86	87-89	90-93	94-100
F	D-	D	D+	C-	C	C+	B-	B	B+	A-	A

PROGRAMMING ASSIGNMENTS



Programming assignments will be completed with Python 3. Assignments may be worked in pairs or individually.

Late assignment policy:

- Assignments may be turned in up to two days late for a 20% grade reduction.
- Assignments over two days late will not be accepted. Contact me if there are extenuating circumstances.

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



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.

COURSE SCHEDULE

*Tentative. Subject to change at instructor's discretion.

8/31 Introduction

CLASSIFICATION

9/5 Data Preprocessing [Ch 1-3]
9/7 Decision Trees [Ch 4.1-4.3]
9/12 Overfitting [Ch 4.4]
9/14 Nearest Neighbor [Ch 5.2]
9/15 **Program 0 Due**
9/19 Naive Bayes [Ch 5.3]
9/21 Evaluating Classifiers [Ch 4.5-4.6, 5.7-5.8]
9/26 Ensemble Methods [Ch 5.6]
9/28 Exam Review / **Program 1 Due on 9/29**
10/3 **EXAM 1**
10/5 SVMs [Ch 5.5]

CLUSTERING

10/10 Clustering & K-means [Ch 8.1-8.2]
10/12 Hierarchical Clustering [Ch 8.3]
10/17 Density-based & Fuzzy Clustering [8.4, 9.2]
10/19 Evaluating Clusters [Ch 8.5]
10/24 Anomaly Detection [Ch 10]
10/26 Exam Review / **Program 2 Due on 10/27**
10/31 **EXAM 2**

ASSOCIATION ANALYSIS

11/2 Apriori [Ch 6.1-6.5]
11/7 FP-Growth [Ch 6.6]
11/9 Compact Itemset Representation
11/14 Sequential Patterns [Ch 7.4]
11/16 Graphs & Subgraphs [Ch 7.5]
11/21 Evaluating Association Patterns [6.7-6.8]
11/23 (No class - Thanksgiving Day)
11/28 Infrequent Patterns [Ch 7.6]
11/30 Guest Speaker (tentative) / **Program 3 Due**
12/5 Exam Review
12/7 **EXAM 3**

ANONYMOUS FEEDBACK

Anonymous feedback can be provided to the instructor at any time via

Canvas -> Quizzes -> Anonymous Feedback

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

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.