The University of Texas at Austin
Department of Civil, Architectural & Environmental Engineering

CE 395R 5-Data Mining – Spring 2018

UNIQUE NUMBER: 14920

INSTRUCTOR: Dr. Carlos H. Caldas
ECJ 5.210
Phone: 471-6014 Fax: 471-3191
e-mail: caldas@mail.utexas.edu
http://www.ce.utexas.edu/prof/caldas/

MEETINGS: Fridays – 10:00AM to 1:00PM
Please refer to class schedule for details
Room: ECJ 2.210

OFFICE HOURS: Tuesdays and Thursdays – 11:00AM to 12:00PM

I encourage students to come see me to address any questions or concerns about the course material or other issues. I have an open door policy – if my office door is open, I will see students without an appointment. If I am busy, we will schedule a convenient time for both of us. If you are unable to come to my office, the next best method to get in touch with me is by phone. I will respond to all phone messages left M-F by 5 pm, within 24 hours. Because of the potential for communication problems, I would prefer that you don't rely on e-mail as our primary method of communication. I will respond to most emails, but owing to problems that can occur in sending and receiving electronic messages, it is much better to come to my office or call me.

WEB PAGE: You will find the online materials for this course at the Canvas site at: http://canvas.utexas.edu/

COURSE OBJECTIVES: This course will discuss data mining (DM) and knowledge discovery in databases (KDD). The class will involve lectures, lab sessions, class discussions, homework assignments, group projects, and student presentations. By taking this class you will be able to:

(1) Understand the fundamentals of data mining and knowledge discovery in databases.
(2) Demonstrate how knowledge discovery in databases can be used to support construction engineering and project management.
(3) Apply DM/KDD techniques for data classification, prediction, clustering, and mining association rules.
(4) Recognize the design, analysis, and implementation issues for DM/KDD techniques in civil engineering.
(5) Analyze, evaluate, and recommend DM/KDD systems for engineers, construction owners, contractors, and/or project managers.

Course objectives will be refined as the course evolves. Each student will therefore be given the opportunity to influence the direction of the course.

REFERENCE MATERIAL:

- Tan, P., Steinbach, M., Karpatne, A., and Kumar, V. (2018) *Introduction to Data Mining*, 2nd edition, Addison-Wesley, ISBN: 13: 978-0-13-312890-1
- Witten, I., Frank, E., Hall, M., and Paul, C. (2016) *Data Mining: Practical Machine Learning Tools and Techniques*, 4th edition, Morgan Kaufmann, ISBN: 978-0128042915.

TOPICS:

DATA MINING AND KNOWLEDGE DISCOVERY IN DATABASES

- Introduction and Overview
- Basic Concepts

ENGINEERING DATA

- Data Types
- Data Preprocessing
- Data Exploration
- Visualization

DATA MINIG TASKS

- Classification
- Prediction
- Association Rules
- Clustering

MINING COMPLEX DATA TYPES

TRENDS IN DATA MINING

ENGINEERING AND PROJECT MANAGEMENT APPLICATIONS

NOTE ON CONTENT OF THE COURSE

The course is necessarily focused, and therefore selected material is left out. Students may pick up a greater depth as necessary in related courses in MIS, CS, and EE in topics such as artificial intelligence, machine learning, data mining, and pattern recognition.

SOFTWARE DEMONSTRATIONS

In-class demonstrations of software tools will be presented. They are intended to provide students with first introductions to the tools and give them a "jump start", not to fully train students on the use of tools. Students will require self-study before they will be able to work effectively with the tools.

GRADING:

Grade components will be weighted as follows in the computation of the final course grade:

Participation	5%
Homework Assignments	25%
Exam	30%
Semester Project	40%

COURSE LETTER GRADES:

The correspondence of letter grade to numerical grade is:

A:	grade $\geq$ 94	D+:	$67 \leq$ grade < 70
A-:	$90 \leq$ grade < 94	D:	$64 \leq$ grade < 67
B+:	$87 \leq$ grade < 90	D-:	$60 \leq$ grade < 64
B:	$84 \leq$ grade < 87	F:	grade < 60
B-:	$80 \leq$ grade < 84		
C+:	$77 \leq$ grade < 80		
C:	$74 \leq$ grade < 77		
C-:	$70 \leq$ grade < 74		

The instructor reserves the right to adjust letter grades, upward only, based on individual attendance and class participation if the numerical grade warrants such consideration.

COURSE/INSTRUCTOR EVALUATIONS:

An evaluation of the course and instructor will be conducted at the end of the semester using the approved UT Course/Instructor evaluation forms.

POLICIES:

Exams and Homework Assignments:

The exam is closed book, closed notes. The exam will include material covered in reading assignments and class discussions. Exam make-ups will be given only in the event of a verified emergency or doctor-verified sickness. No final exam will be given.

The student is responsible for all reading assignments and class handouts whether or not covered in class or listed on the syllabus.

There will be individual and group homework assignments. Their due dates will be posted on the course website. All assignments **are due at the beginning of the period assigned** and those turned in late will count off **10% per day**. (no exceptions!--except those listed for the test make-up).

There will be a semester project. More information concerning the semester project will be given at a later date.

My goal is to grade all exams and homework assignments within two weeks.

Homework and exams will not be returned to the student. The homework and exams solutions will be reviewed in class. You can view your homework and exams during office hours.

Class Participation and Attendance:

It is important that you are familiar with the course material as the course evolves. Your ability to answer questions and discuss the material will be part of the overall participation evaluation. Therefore, you should review class material ahead of time. Regular attendance is expected and encouraged. **Your attendance will be used to evaluate your participation grade**. I consider a student missing more than one week of class lectures without excuse to be a serious participation problem. In some cases, I will petition the Office of Student Affairs to drop students from the course who have excessive absences and may withhold the entire participation grade at my discretion for participation problems. Each student is responsible for all material and administrative instructions given during the lecture period.

A student who misses classes or other required activities, including examinations, for the observance of a religious holy day should inform the instructor as far in advance of the absence as possible, so that arrangements can be made to complete an assignment within a reasonable time after the absence.

Personal Problems:

If you have illness or personal problems that will affect your performance during the course of the semester, please let me know as soon as possible. "After the fact" provides little protection unless there are extreme circumstances. I have an answering machine, a fax machine, and an e-mail address if you need to get in touch with me after hours. Do not hesitate to use them.

Scholastic Dishonesty:

IMPORTANT! Students who violate University rules on scholastic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on scholastic dishonesty will be strictly enforced. For further information, visit the Student Judicial Services web site http://deanofstudents.utexas.edu/sjs/, and the General Information Catalog information at http://registrar.utexas.edu/catalogs/gi07-08/app/appc03.html#Chapter-11-Student-Discipline-and-Conduct.

Students with Disabilities:

The University of Texas at Austin provides, upon request, appropriate academic accommodations for qualified students with disabilities. For more information, contact the Division of Diversity and Community Engagement, Services for Students with Disabilities, 512-471-6259 (Videophone: 512-410-6644) or http://diversity.utexas.edu/disability/.

Privacy – Web Based Class Sites:

Web-based, password-protected class sites will be associated with all academic courses taught at the University. Syllabi, handouts, assignments and other resources are types of information that may be available within these sites. Site activities could include exchanging e-mail, engaging in class discussions and chats, and exchanging files. In addition, electronic class rosters will be a component of the sites. Students who do not want their names included in these electronic class rosters must restrict their directory information in the Office of the Registrar, Main Building, Room 1. For information on

restricting directory information, see the General Information Catalog or go to: http://registrar.utexas.edu/catalogs/gi07-08/app/appc02.html#Chapter-9-Educational-Records.

Dropping the Class:

Undergraduate Students: From the 1st through the 12th class day, an undergraduate student can drop a course via the web and receive a refund, if eligible. From the 13th through the university's academic drop deadline, a student may Q drop a course with approval from the Dean, and departmental advisor.

Graduate Students: From the 1st through the 4th class day, graduate students can drop a course via the web and receive a refund. During the 5th through 12th class day, graduate students must initiate drops in the department that offers the course and receive a refund. After the 12th class day, no refund is given. No class can be added after the 12th class day. From the 13th through the 20th class day, an automatic Q is assigned with approval from the Graduate Advisor and the Graduate Dean. From the 21st class day through the last class day, graduate students can drop a class with permission from the instructor, Graduate Advisor, and the Graduate Dean. Students with 20-hr/week GRA/TA appointment or a fellowship may not drop below 9 hours.

Computer Usage:

Students are expected to be proficient on a personal computer and to be able to use CAD, word processing, and spreadsheet programs. Familiarity with the Civil Engineering Learning Resources Center (LRC) is assumed. The web-based UT Canvas system will be used extensively to coordinate class assignments and disseminate course information, including class notes.

Recommendations Regarding Emergency Evacuation from the Office of Campus Safety and Security
Contact Info: 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 (across the bridge).
- 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 Dept., The University of Texas at Austin Police Dept., or Fire Prevention Services office.
- Behavior Concerns Advice Line (BCAL) 512-232-5050. For more information visit the BCAL website: http://www.utexas.edu/safety/bcal/
- Link to information regarding emergency evacuation routes and emergency procedures can be found at: www.utexas.edu/emergency

All other university policies not explicitly included on this syllabus can be found on the General Information Catalog: http://catalog.utexas.edu/general-information/

IMPORTANT DATES:

No classes on March 16th due to Spring Break. January 19th is the last day graduate students may register and pay tuition without the approval of the graduate dean. January 31st is the last day to drop a class for a possible refund. April 2nd the last day a graduate student may change registration in a class to or from a credit/no credit basis. April 30th is the last day a graduate student may, with the required approvals, drop a class or withdraw from the University.

Midterm exam: April 13th.

SCHEDULE:

Note: The course schedule is subject to changes. Any changes in the course schedule will be communicated in advance and posted in the course Canvas web page.

Week	Date	Topic	Reference Material	Assignment Due
1	Jan 19	Introduction; Data Mining Overview	TSK Ch. 1	
2	Jan 26	Data Types; Data Preprocessing; Data Exploration	TSK Ch. 2	
3	Feb 2	Data Exploration	TSK Ch. 2 & 3	
4	Feb 9	Data Exploration; Data Visualization	TSK Ch. 3	
5	Feb 16	Classification	TSK Ch. 4	
6	Feb 23	Classification	TSK Ch. 5	
7	Mar 2	Prediction	TSK Ch. 5	
8	Mar 9	Cluster Analysis	TSK Ch. 8	
9	Mar 16	SPRING BREAK		
10	Mar 23	Mining Association Rules	TSK Ch. 6	Semester Project Deliverable 1
11	Mar 30	Mining Association Rules	TSK Ch. 7	
12	Apr 6	Mining Complex Data Types; Trends		
13	Apr 13	EXAM		
14	Apr 20	Mining Complex Data Types; Trends		
15	Apr 27	Semester Project Presentations: Final	Project Presentations	
16	May 4	Semester Project Presentations: Final	Project Presentations	Semester Project Deliverable 2