

Instructor Contact Information

Instructor: Michael Hahsler
Office: 311 Caruth
Phone: (214) 768-8878
E-mail: mhahsler@lyle.smu.edu
Office hours: Tue 12:30-2:00pm, Thu 9:00-10:30am or by appointment

Course Information

Lecture: TTh 2:00-3:20pm
Course website: <http://michael.hahsler.net/SMU/EMIS7331/>
For further information, assignments and assignment submission:
<http://canvas.smu.edu/>

Course Description

Analytics is based on collecting, managing, exploring and acting on large amounts of data and has become a source of competitive advantage for many organizations. This course provides an overview of descriptive analytics and introduces major data-mining techniques (classification, association analysis and cluster analysis) used in predictive analytics. All material covered will be reinforced through hands-on experience using state-of-the art tools (currently R) to design and execute data mining processes.

Prerequisites:

- Introduction to programming: CSE 1342 (any programming language)
- Probability and statistics: CSE 4340/EMIS 3340/STAT 4340 or CSE 7370/EMIS 7370
- Databases: EMIS 3309 or CSE 3330 (undergraduates only)

Learning Outcomes

1. Students will be able to explain and use the mining process for descriptive and predictive analytics.
2. Students will be able to use R for data preparation, data exploration and analysis, and predictive modeling.
3. Students will understand and be able to apply the core data mining methods of
 - Classification
 - Association Analysis
 - Cluster Analysis
4. Students will be able to conduct a complete data mining project including research, data preparation, and reporting the results.

Unofficial Learning Outcome

The course consists of four major projects, where you will be asked to answer given questions using provided data and the tools covered in class. However, the aim is that by the end of this course you can

1. formulate your own questions and clearly explain why answers to these questions are significant,
2. communicate the value of these answers to stake-holders, and
3. identify and use the appropriate data and techniques to obtain the answers.

Textbook

Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, University of Minnesota, University of Minnesota, Addison Wesley, 2006. **Note: You may use any edition of this book.**

We will cover chapters 1-9 in this course.

Course Topics and Class Outline

Session	Date	Lecture	Reading	Project
Week 1	1/23 & 25	Intro + Analytics and the Data Mining Process	Ch 1	
Week 2	1/30 & 2/1	Data	Ch 2	
Week 3	2/6 & 8	Data	Ch 3	
Week 4	2/13 & 15	Data		
Week 5	2/20 & 22	Data, Clustering	Ch 8	P1 due on 2/22
Week 6	2/27 & 3/1	Clustering		
Week 7	3/6 & 8	Clustering	Ch 9	
Week 8	3/13 & 15	Spring break		
Week 9	3/20 & 22	Classification	Ch 8	P2 due on 3/22
Week 10	3/27 & 29	Classification		
Week 11	4/3 & 5	Classification	Ch 9	
Week 12	4/10 & 12	Classification		
Week 13	4/17 & 19	Association Analysis	Ch 6	P3 due 4/19
Week 14	4/24 & 26	Association Analysis		
Week 15	5/1 & 3	Association Analysis, Advanced Topics	Web Site	P4 due on 5/10

Course Work and Grading

It is expected that each student will keep up with the reading as outlined above. Additional materials may be referenced in class as needed. The course grade will be determined based on the four projects and peer evaluation.

Assessment	Percentage
Projects*	23% each
Peer Evaluation	2% each
Online Participation (Canvas discussion board)	Up to 1% bonus per project

Note: There is no midterm or final examination! Your entire grade depends on your performance on the four projects. The projects are hard work and designed to take several weeks each to complete. So, start early and start strong!

* Students enrolled in the graduate section will have additional project requirements.

Projects

You will work in **teams of two students** (distance students may elect to work alone) on four projects throughout the semester. The result of each project is a detailed project report (following the categories of the CRISP-DM framework). You may share code, comments and ideas between teams (using the discussion board on canvas), **but each team needs to write their own report independently!** The report needs to be submitted via Canvas before the due date. Late submissions of up to 3 days are possible, but you will lose 10% per day.

You can expect a B (up to 90%) if you address all questions in the project description adequately. For a better grade you need to present exceptional work (e.g., asking and answering your own questions, in-depth explanation why one method works better than another, developing special preprocessing, insightful visualization, using and explaining a method not covered in class).

Peer Evaluation

For each project you will be assigned two reports from other students. The evaluation will be done via Canvas and the other student will not see who evaluated their report. You will be graded on the completeness and quality of the reviews you write, not on the reviews your work receives!

Online Participation

Frequent participation on the classes discussion board (on Canvas) will earn you up to 1% bonus per project.

Attendance Policy

Students are expected to attend class regularly. **If a student is absent from class, it is that student's responsibility to make arrangements with the professor to make up any work missed or to ensure that assignments are submitted on time or early.** Late assignments will not be accepted except in extreme instances. Any assignments that will be missed (including those due to university-sanctioned events) must be completed before the due date. This includes quizzes and homework assignments.

Academic Ethics and Collaboration

You may work together to get ideas and write code, but you are expected to create, edit and print your own assignments. If you submit work which is not completely yours then you will receive a 0 on the particular assignment. In severe cases, you will receive an F in the course and may be brought in front of the SMU Honor Council. It is your responsibility to know and understand the University's Honor Code and the expectations for collaboration in this course.

Additional Information

- **Disability Accommodations:** Students needing academic accommodations for a disability must first register with Disability Accommodations & Success Strategies (DASS). Students can call 214-768-1470 or visit <http://www.smu.edu/Provost/ALEC/DASS> to begin the process. Once registered, students should then schedule an appointment with the professor as early in the semester as possible, present a DASS Accommodation Letter, and make appropriate arrangements. Please note that accommodations are not retroactive and require advance notice to implement.
- **Religious Observance:** Religiously observant students wishing to be absent on holidays that require missing class should notify their professors in writing at the beginning of the semester, and should discuss with them, in advance, acceptable ways of making up any work missed because of the absence. (See University Policy No. 1.9.)
- **Excused Absences for University Extracurricular Activities:** Students participating in an officially sanctioned, scheduled University extracurricular activity should be given the opportunity to make up class assignments or other graded assignments missed as a result of their participation. It is the responsibility of the student to make arrangements with the instructor prior to any missed scheduled examination or other missed assignment for making up the work. (University Undergraduate Catalogue)