THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

DATA MINING
FOR BUSINESS INTELLIGENCE

PROFESSOR MAYTAL SAAR-TSECHANSKY

Data Mining For Business Intelligence: MIS 382N.9/MKT 382

Professor Maytal Saar-Tsechansky

This course provides a comprehensive introduction to data mining problems and tools to enhance managerial decision making at all levels of the organization and across business units. We discuss scenarios from a variety of business disciplines, including the use of data mining to support customer relationship management (CRM) decisions, decisions in the entertainment industry, finance, and professional sports teams.

The three main goals of the course are to enable students to:
1. **Approach business problems data-analytically** by identifying opportunities to derive business value using data mining.
2. **Interact competently on the topic** of data-driven business intelligence. Know the basics of data mining techniques and how they can be applied to interact effectively with CTOs, expert data miners, and business analysts. This competence will also allow you to envision data-mining opportunities.
3. **Acquire some hands-on experience** so as to follow up on ideas or opportunities that present themselves.

Reading Materials and Resources

1. Reading materials posted on Blackboard or distributed in class
2. Textbook: Data Mining Techniques, Third Edition by Michael Berry and Gordon Linoff Wiley, 2004 ISBN: 0-471-47064-3

Software: WEKA (award-winning, open source software tool)

Course Requirements and Grading

Style
This is a lecture-style course, however student participation is important. Students are required to be prepared and read the material before class. Students are required to attend all sessions and discuss with the instructor any absence from class. We will also have several guest speakers from a variety of industries who will discuss how they apply data mining techniques to boost business performance.

Assignments
You will hand in 6-8 assignments which address the materials discussed in class as well as help you develop hands-on experience analyzing business data with a user-friendly and easy-to-use data mining software tool.
Assignments will be announced in class and be posted on blackboard. The due date of each assignment will be a week from the day in which it is announced in class. The due date for each assignment will also be noted next to the assignment on Blackboard. It is the student's responsibility to review the class's Blackboard site for homework and their due dates.
50% of the assignments will be graded.

Term Project
There is no final exam. Throughout the second half of the semester students will be working in teams (preferably, around 3-4 students per team) to analyze and develop a solution to business problems. The students will use real business data to which they will apply data mining techniques. Teams are required to hand in a brief report (95% of project grade) and prepare a short class presentation of their work (5% of project grade). Time permitted, a class discussion will follow the presentations.

> **Deliverables**:
> Each team will hand in a brief report (95%) and prepare a short presentation (5%) of their work.
> Each team member will also provide feedback on the contribution of each of the other team members. Feedback must be provided via email to the instructor by the last day of class. Your team evaluation may raise, decrease or not affect your project grade.

Late assignments
Assignments are due prior to the start of the lecture on the due date. **Please turn in your assignment early if there is any uncertainty about your ability to turn it in on the due date**. Assignments up to one week late will have their grade reduced by 50%. After one week, late assignments will receive no credit. There will be no exceptions.

Midterms
There will be 2 midterms during the course of the semester. Please review midterm dates in the schedule below. Midterms will be brief and their objective is to review key concepts introduced in the recent modules. Format: each student will answer the quiz individually. Students will then be divided into groups to discuss and retake the quiz as a group. The group discussion will follow by a review of the correct responses. A correct response by the group will add up to 10 points. Even if you answered the individual quiz correctly, you will benefit from the extra points. Thus group discussion can only help all members of the group.

Missed midterms
If you miss a midterm without excuse, you will receive zero points. Valid excuses for missing a midterm are, for example, (documented) illness, death of a family member, or a meeting with the president. These excuses will have to be documented. To make up points for an excused absence, you will have an oral exam (15-20 min) with me. You will not receive the team bonus points for the missed midterm.

Grade breakdown:
1. Involvement : 10%
2. Midterms : 40%
3. Assignments: 15%
4. Team project: 35%

Office Hours

Professor Saar-Tsechansky: Thursday 1-2pm, and by appointment.CBA 5.230.

Sam Blazek (Teaching Assistant):
Samuel.Blazek@phd.mccombs.utexas.edu

Both the TAs and myself are available during posted office hours as well as at other times by appointment. Do not hesitate to request an appointment if you cannot make it to the posted office hours. The most effective way to request an appointment for office hours is to suggest several times that work for you.

Communication, Text, etc.

The Blackboard site for this course will contain lecture notes, reading materials, assignments, and late breaking news. It is accessible via: www.**utexas**.edu/cc/**blackboard/**
Assignments will be announced in class and students are responsible to check Blackboard after each class to download assignments and for information on assignments due dates.

Email policy

Emails to me or the TAs should be restricted to organizational issues, such as requests for appointments, questions about course organization, etc. For all other issues, please see us in person. Specifically, we **will not discuss technical issues related to homeworks or projects per email**. Technical issues are questions concerning how to approach a particular problem, whether a particular solution is correct, or how to use the software. It is fine to inquire per email if you suspect that a problem set has a typo or if you find the wording of a problem set ambiguous.

Email: maytal@mail.utexas.edu ß□Begin subject: [DM GRAD]…

McCombs Classroom Professionalism Policy

The highest professional standards are expected of all members of the McCombs community. The collective class reputation and the value of the Texas MBA experience hinges on this.

Faculty are expected to be professional and prepared to deliver value for each and every class session. Students are expected to be professional in all respects.

The Texas MBA classroom experience is enhanced when:

- **Students arrive on time.** On time arrival ensures that classes are able to start and finish at the scheduled time. On time arrival shows respect for both fellow students and faculty and it enhances learning by reducing avoidable distractions.
- **Students display their name cards.** This permits fellow students and faculty to learn names, enhancing opportunities for community building and evaluation of in-class contributions.
- **Students minimize unscheduled personal breaks.** The learning environment improves when disruptions are limited.
- **Students are fully prepared for each class.** You will learn most from this class if you work and submit homework on time, keep up with the content introduced in each session, and come prepared to class.
- **Students respect the views and opinions of their colleagues.** Disagreement and debate are encouraged. Intolerance for the views of others is unacceptable.
- **Laptops are closed and put away.** Except for session in which we will use the WEKA software tools, I request that laptops will remain closed and put away. When students are surfing the web, responding to e-mail, instant messaging each other, and otherwise not devoting their full attention to the topic at hand they are doing themselves and their peers a major disservice. There are often cases where learning is enhanced by the use of laptops in class. In sessions when WEKA will be used, I expected students to behaved professionally.
- **Phones and wireless devices are turned off.** When a need to communicate with someone outside of class exists (e.g., for some medical need) please inform the professor prior to class.

Your professionalism and activity in class contributes to your success in attracting the best faculty and future students to this program.

Academic Dishonesty

Please keep in mind the McCombs Honor System.
http://mba.mccombs.utexas.edu/students/academics/honor/index.asp

Students with Disabilities

Upon request, the University of Texas at Austin provides appropriate academic accommodations for qualified students with disabilities. Services for Students with Disabilities (SSD) is housed in the Office of the Dean of Students, located on the fourth floor of the Student Services Building. Information on how to register, downloadable forms, including guidelines for documentation, accommodation request letters, and releases of information are available online at http://deanofstudents.utexas.edu/ssd/index.php. Please do not hesitate to contact SSD at (512) 471-6259, VP: (512) 232-2937 or via e-mail if you have any questions.

Tentative Course Schedule

Date	Topic
1/14	Data mining, Machine Learning and Artificial Intelligence
1/16	Introduction (Cont'd) Fundamental concepts and definitions: The data mining process Data mining predictive and descriptive tasks Additional reading (BB): The KDD process for extracting useful knowledge from volumes of data. Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth. Communications of the ACM. Volume 39, Issue 11 (November 1996). ACM Press New York, NY, USA (http://citeseer.ist.psu.edu/fayyad96kdd.html) Competing on Analytics by Thomas Davenport. Don Cohen, and Al Jacobson. May 2005 (http://www.babsonknowledge.org/analytics.pdf)
1/21	Classification: Recursive partitioning & Decision Trees
1/23	Classification: Recursive partitioning & Decision Trees
1/28	Model Evaluation
1/30	Model Evaluation
2/4	Model Evaluation
2/6	WEKA lab session : Introduction to WEKA
2/11	WEKA lab session : Introduction to WEKA
2/13	Recommender systems: K-Nearest Neighbor Algorithm & Collaborative Filtering
2/18	Recommender systems: K-Nearest Neighbor Algorithm & Collaborative Filtering
2/20	**Midterm**
2/25	Recommender systems: Association rules, and PageRank
2/27	WEKA lab session: Association rules Basketball memorabilia: Decision making using data-driven business intelligence Evaluating decision making strategies
3/4, 3/6	**Plus Program** **Optional in-class consultation on team projects**
3/11, 3/13	**Spring Break**

Date	Topic
3/18	Clustering & Segmentation
3/20	**Guest Speaker: Dr. Jeff Castura** Jeff holds a Bachelor of Science in Engineering Physics from Queen's University, Canada and a Master of Applied Science in Electrical Engineering from the University of Toronto, Canada. He received his Ph.D. in Electrical Engineering from the University of Ottawa, Canada in 2008. His industry experience includes ASIC and semiconductor design, algorithm development and system architecture for wireless communication systems. Jeff is currently with RGM Advisors LLC, where he applies statistical and machine learning predictive modeling techniques to analyze and predict financial time-series.
3/25	Clustering & Segmentation: WEKA lab session, and GE Case
3/27	Basketball case, hands-on session Decision making using data-driven business intelligence Evaluating decision making strategies
4/1	Text mining and information retrieval: Bayesian learning with applications to spam filtering: conditional probability, Bayes rule, Naïve Bayes classifier. Bag-of-Words representation, analysis of blogs, news stories, sentiment elicitation.
4/3	Text Mining (Cont'd) and applications (See supplement readings on Blackboard) Time permitted: In-class work on term project
4/8	**Midterm**
4/10	Artificial Neural Networks
4/15	**Guest Speaker: Tal Tvertsky, Apple (Data Mining Group)**
4/17	Genetic Algorithms
4/22	Genetic Algorithms In-class work on term projects (bring laptop)
4/24	In-class advice on term projects
4/29	**Term project is due** Team projects - presentations and discussion
5/1	Team projects - presentations and discussion