THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

BUSINESS INTELLIGENCE WITH DATA MINING

PROFESSOR MAYTAL SAAR-TSECHANSKY

Data Mining: MIS 373/MKT 372
Professor Maytal Saar-Tsechansky

"***For every leader in the company, not just for me, there are decisions that can be made by analysis. These are the best kinds of decisions. They're fact-based decision***s." Amazon's CEO, Jeff Bezos.

In the January 2006 *Harvard Business Review* issue, Professor Thomas Davenport and colleagues document the emergence of a new form of **competition based on the extensive use of analytics**, data, and fact-based decision making. In virtually every industry, the authors found the competitive strategies organizations are employing today rely extensively on data analysis to predict the consequences of alternative courses of action, and to guide executive decision making, more generally. Extensive interviews with executives from successful firms find that companies today require **decision makers** who **understand the value of analytics**, can **identify opportunities** and **know how best to apply** data analytics to enhance business performance. The spreading of analytical competition spans industries—from consumer finance to retailing to travel and entertainment to consumer good, and even professional sports teams.

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 from 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. Textbook: Data Mining Techniques, Third Edition by Michael Berry and Gordon Linoff Wiley, 2004 ISBN: 0-471-47064-3
2. Additional reading materials will be available on blackboard.

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.

Individual assignments
Individual assignments address the materials discussed in class as well as aim to help you develop hands-on experience analyzing business data with a data mining software tool. Assignments will be announced in class and be posted on blackboard. Students are responsible to know when assignments are due. The due date of each assignment will be a week from the day in which it will be announced in class. The due date will be also noted on Blackboard next to each assignment.

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. Legitimate reasons for an inability to submit an assignment on time must be supported by appropriate documentation. There will be no exceptions.

Quizzes
There will be 4 quizzes during the course of the semester. Please review Quiz dates in the schedule below. Quizzes 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.
There will be a 5th **optional** quiz, which is not mandatory. Students taking the 5th quiz can take it in order to improve their average performance in the quizzes. The grade on the 5th quiz cannot replace a low grade or a zero for a missed quiz, but will be added to the average.

Missed quizzes
If you miss a quiz without excuse, you will receive zero points. Valid excuses for missing a quiz are, for example, illness, death of a family member, or a meeting with the president. A job interviews is not a valid excuse -- please be sure not to schedule interviews on these dates. In addition <u>excuses will have to be documented</u>. To make up points for an excused absence, you will have an oral exam. You will not receive the team bonus points for the missed quiz.

Team project
There will be a final term project in which teams can chose between developing a proposal for a data mining project to address a business problem, or a hands-on data project in which the team will address a business problem by applying data mining techniques to real business data.

> **<u>Deliverables</u>**:
> Each team will hand in a brief report (85%) and prepare a short presentation (15%) of their work.
> Each team member will also provide feedback on the contribution of each of the team members. Feedback must be provided via email to the instructor by the last day of class. Your grade for the team project may be raised, decrease or not be affected by your teammates' feedback, depending how your performed relative to the other team members.

Attendance:
Attendance will be taken in each class. Any absence must be supported by a document, such as from a doctor. Interviews, other class projects, etc. will not be accepted as legitimate reason to miss a class. There will be no exceptions to this policy.

Grade breakdown:
1. Involvement : Includes attendances, interest and effort: 10%
2. Assignments : 10%
3. Quizzes: 50%
4. Group term project (teams): 30%

Course Materials

All course-related materials, such as handouts, announcements, slides, etc., will be posted on Blackboard (http://courses.utexas.edu).

Office Hours

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

Sam Blazek (Teaching Assistant):
Email: sam.blazek@gmail.com.

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.
Please note that I will usually not be able to make appointments before 9:00am or after 5:00pm.

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 quizzes or homeworks 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 Ok 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 UNDERGRAD]…

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 McCombs BBA program 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 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.** 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.

- **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.

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	Introduction to the course. Introduction to data mining. What is data mining? Why now?
1/16	Fundamental concepts and definitions: The data mining process Data mining predictive and descriptive tasks Supplement reading : 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	Finalize classification trees, inference with trees, inference with missing values.
1/30	Model Evaluation: Classification accuracy rate, and cost sensitive evaluation metrics
2/4	**Review for Quiz 1**
2/6	**Quiz 1: Introduction and classification trees**
2/11	Model Evaluation
2/13	Model Evaluation
2/18	WEKA lab session
2/20	WEKA lab session
2/25	**Quiz 2 : Model evaluation**
2/27	Recommender Systems: Personalized recommendations, Person-to-person recommendations and KNN algorithm
3/4	Recommender Systems: Personalized recommendations, Person-to-person recommendations and KNN algorithm .
3/6	Recommender systems: Person-to-person, item-to-item, association rules, sequential patterns, PageRank.
3/10, 3/13	**Spring Break**

Date	Topic
3/18	WEKA Lab Session: Association rules Basketball memorabilia investment case (time permitting)
3/20	**Quiz 3:** Recommender Systems: Content-based recommendations and collaborative filtering
3/25	Decision making using data-driven business intelligence Evaluating decision making strategies
3/27	WEKA Lab Session: Basketball memorabilia investment case
4/1	Clustering/segmentation analysis **<Intermediate report on term project is due at the beginning of class>**
4/3	Clustering, GE Case
4/8	WEKA Lab Session: Clustering NBA players
4/10	**Quiz 4** : Item-to-item vs. person-to-person recommender systems, network-based recommendations (Page-rank), clustering, and lessons from the Basketball memorabilia investment case.
4/15	No class (Passover)
4/17	Text mining and information retrieval: Bayesian learning with applications to spam filtering: conditional probability, Bayes rule, Naïve Bayes classifier
4/22	**Optional Make-up Quiz**
4/24	In class work on term projects
4/30	**Term project report is due in class** Team projects - presentations and discussion
5/1	**Feedback on team members' contributions is due** Team projects - presentations and discussion