Boston University School of Management
Information Systems Department
IS 841 – Advanced Business Analytics: Data Mining

Fall 2014

Instructor: Nachiketa Sahoo
Office: 627 SMG Building
Email: nachi@bu.edu
Office Hours: Tuesday 4-6pm or by appointment
Class Meeting Day/Time: Tuesday 6-9pm
Class Room: 406 SMG Building

What to do before our first class meeting (Tuesday September 2):

1. Order/purchase the textbook (see below)

2. Complete the registration and installation steps for accessing SAS On Demand: Enterprise Miner (see below). We will troubleshoot and answer any questions in class.

Course Description and Objectives:

The widespread proliferation of IT-influenced economic activity leaves behind a rich trail of micro-level data about consumer, supplier and competitor preferences. This has led to the emergence of a new form of competition based on the extensive use of analytics, experimentation, and fact-based decision making. In virtually every industry 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. This course provides a hands-on introduction to the concepts, methods and processes of business analytics. We will learn how to obtain and draw business inferences from data by asking the right questions and using the appropriate tools. Topics to be covered include data preparation, data mining, text mining, recommender systems as well as the overall process of using analytics to solve business problems, its organizational implications and pitfalls. We will work with real world business data and analytics software. Where possible, cases will used to motivate the topic being covered.

Prerequisites:

If you have taken a basic database, statistics class (even in your undergraduate work) and/or have some relevant prior work experience you can register for this course. If you have any questions please contact nachi@bu.edu

Textbook/Readings:

Data Mining Techniques: For Marketing, Sales and Customer Relationship Management, **THIRD EDITION (2011)**; by Michael Berry and Gordon Linoff.

Available on Amazon and other online bookstores. http://www.amazon.com/dp/0470650931/
The book should also be available at the Barnes-and-Nobles bookstore in Kenmore square. Please make sure you purchase the latest (3rd) edition of this book.

Additional readings will be made available online via SMGTools under Resources -> Slides and Readings. Readings are organized by session and labeled as "required" or "recommended."

How to prepare for each class: Read and prepare to discuss any assigned cases/HBR articles etc. **before** class. Assigned textbook chapters and other technical readings are best done **after** class.

Software:

We will use **SAS Enterprise Miner** as our data-mining tool. We will also introduce **NodeXL**, a social network analysis tool.

SAS EM is a high-end professional tool that you are likely to encounter in your jobs. SAS offers a cloud-based version of the tool (SAS On Demand for Academics) for free to qualified academic institutions. It runs on most Java enabled web browsers.

To obtain access to this tool you need to register with SAS. You can register for this course by following the step-by-step guide at http://support.sas.com/ondemand/manuals/EnterpriseMinerStudent.pdf
You'll need the following course link in the second step of the registration:
https://odamid.oda.sas.com/SASODAControlCenter/enroll.html?enroll=4eee367a-9ae2-4266-8b67-f30722e1bd08

SAS mentions that the cloud-based tool is only guaranteed to work under Windows. Although, last years' class was able to use it without problems from Mac and Linux laptops, it's useful to remember that SAS does not provide any support for Mac/Linux users. So, those of you with Macs may want to use a Windows emulator. Since this is a data intensive application, it is highly recommended that you use an **Ethernet cable** instead of wireless connection for best user experience.

It will be helpful if you complete the SAS registration and installation steps before our first class session. In any event, we will go through the registration and installation process together during the first day of classes.

NodeXL is a free, open-source template for Microsoft® Excel® 2007, 2010, and 2013 that makes it easy to explore network graphs. I will provide installation details later on during the semester.

Course Deliverables and Grading:

Deliverable	**Weight**
Two write-ups (see below)	10%
Midterm Examination	25%
Final Examination	25%
Team Project	30%
Class Participation	10%

Deliverable Descriptions:

(1) Two 1-2 page write-ups due 9pm on day before the class. Data Mining Opportunity (10%):

For 10 classes I will select a firm (e.g. "Federal Express"). For two of the classes (see SMGTools for specific dates when you will have to turn in your papers) each of you will need to turn in a 1-2 page write-up (**via smgtools assigments, not by email**) **by 9pm of the day before the class** about ONE SPECIFIC

OPPORTUNITY for this company to benefit from data mining. *You should also be prepared to discuss your idea in class for 3-5 minutes. You can also prepare 1-2 (max) slides and bring it to the class. In the past, bringing such slides to class has been found to improve the quality of discussion*.

You should think about and address the following in the write-up:

- Description of the specific opportunity for data mining – do NOT list many ways in which the firm can use data mining. Be specific and choose **one** example which you will clearly state and describe.
- Describe the data that is useful for this opportunity. **Be very specific**, list variable names if you can and provide a simple snapshot (maybe in an Excel table) of some example "rows" in this database.
- Describe how the firm can evaluate the data mining model built. Address this in two ways. First, what "test" data can be used to score the data mining model(s)? Second, is there a way for the firm to assign a monetary value based on the performance of the data mining method?
- Briefly describe how the final data mining model built can be integrated into the appropriate information system (try and identify which systems will use this model and how).
- Try to be creative and **original** (i.e. try to think beyond the most obvious opportunities). *Highly original write-ups (i.e. write-ups that propose **plausible** ideas that no other student proposed) will automatically receive 10% extra credit for the assignment.*

Please note that there is a 2-page maximum limit (11 point or larger font) for your write-up. Late write-ups (i.e. write-ups submitted after 9pm of the day before the class) will be penalized by 5% if 1 hour late or less, 10% if 2 hours or less and 15% if 2 hours or more.

(2) Mid-term and Final Examinations (2x25%=50%).

The exams will be based on class notes, discussion in class and on the relevant readings. Anything discussed in class can be tested in these - unless explicitly excluded - hence presence and participation in class is critical to do well in the exam. The format will be mainly short-answer questions and possibly exercises involving interpreting the results from SAS EM analysis.

(3) Team Project (30%).

In teams of 4-5 students identify a dataset on your own, perform a data mining analysis to answer a plausible question (ideally, but not necessarily, related to business) and summarize the results in a 10-minute presentation/demonstration that will be given on **December 2**. Some examples will be discussed in class.

The following are the detailed project steps/deliverables:

1. I will divide the class into teams. Team assignments will be announced by the second class session.
2. **(5 points) A 2-page project proposal is due on October 7**. The proposal should summarize (a) the problem that motivates the analysis you will conduct, (b) the data set you plan to use, (c) your proposed analysis methodology.
3. **(5 points) A mid-project report, no longer than 2 pages, is due on November 11.** This report should discuss initial results after running preliminary models in SAS EM. You may provide an additional document with tables/figures.
4. **(20 points) The following are due on December 2**:
    a. A brief write-up discussing (a) the problem, (b) the data, (c) your analysis methodology and (d) your analytical findings, conclusions and business recommendations.
    b. A 15-minute presentation with a content and structure similar to the write-up.

c. Your data sets (unless proprietary)
d. A SAS report document showing all analyses performed

Please submit all the documents except the slides by 9pm on 12/1.

Your project grade will based on **(i)** the **novelty of the application,** (ii) your team's **attention to detail** in the analysis performed (i.e. data collection, evaluation methodology, methods etc) and (iii) the **quality of your final presentation and write-up**.

(4) Class Participation

You are expected to attend all class meetings and to be an active participant in class discussion. You are also encouraged to be active on the course's forum that can be accessed through SMGTools course site. The role of the Forum is to provide a platform for posting and discussing interesting stuff you have come across that relates to business analytics and any aspect of the course. Feel free to post articles, cases, visualizations, videos, links to data sets, quotes, or anything else you think might be interesting to the class. I will be doing the same.

Also, there is a support forum on the SMGTools course website where you can ask for help related to any aspect of the class—technical or conceptual. I will monitor the forums and respond in forum. You can also "watch" the forums so that you get notified when there is a new message. Please try to answer a question when you can. By doing this we'll create a shared knowledgebase that can be very helpful as you are trying to figure our how to do certain things, or better understand certain aspects of the course.

The quality (not the quantity!) of your presence in both these forums will be taken into consideration when calculating your class participation grade.

Class Expectations:

Students are expected to adhere to the Boston University code of academic conduct. If you have any concerns or questions about how these policies may apply to this course (or to a specific assignment), please check with your instructor.

COURSE TIMELINE

Session	Date	Topics	Textbook readings*	Write-up due for
1	9/2/2014	Course Introduction; The Data Mining Methodology	Ch 1-4	
2	9/9/2014	Introduction to SAS Enterprise Miner; Predictive Modeling I (Decision Trees)	Ch 5,7	Whole Foods
3	9/16/2014	Predictive Modeling II hands-on exercise	Ch 8,9	American Express (or any credit card company)
4	9/23/2014	Predictive Modeling II (Logistic Regression)		Starwood (or any other hotel chain)
5	9/30/2014	Predictive Modeling III (Neural Networks, Nearest Neighbor approaches)	Ch 12	Any hospital/ Quest diagnostics
6	10/7/2014	Clustering Analysis; Midterm exam review. **Project Proposals Due**	Ch 13, 14	Boston Red Sox (or any sports team)
	10/14/2014	No class		
7	10/21/2014	*1st Examination*		
8	10/28/2014	Association Rules mining.	Ch 15	Google
9	11/4/2014	Variable transformation & reduction; Project Progress Lab I	Ch 19, 20	Amazon.com
10	11/11/2014	Text Mining **Project Progress Report Due**	Ch 21	LinkedIn
11	11/18/2014	Personalization and Recommender Systems		Netflix
12	11/25/2014	Social Network Analysis Guest lecturer : Dylan Walker	Ch 16	Facebook
13	12/2/2014	Project Presentations; Course Wrap up **Project Final Deliverables Due**		
14	12/9/2014	*2nd Examination*		

* Additional readings for most sessions are available on SMGTools. Please check each session's folder under Resources -> Slides and Readings