1. Course Overview

This course will change the way you think about data and its role in business.

Businesses, governments, and society leave behind massive trails of data as a by-product of their activity. Increasingly, decision-makers rely on intelligent systems to analyze these data systematically and assist them in their decision-making. In many cases automating the decision-making process is necessary because of the volume of data and the speed with which new data are generated. This course addresses how technology can be used to connect data to decision-making. The use of real-world examples and cases places data-mining techniques in context and fosters the development of data-analytic thinking, and also illustrates that proper application of data mining techniques is as much an art as it is a science. In addition to the cases, the course features hands-on exercises with data mining software. The course is suitable for those interested in working with and getting the most out of data as well as those interested in understanding data mining from a strategic business perspective.

The goal of this course is three-fold. After taking this course you should:

1. **Approach business problems data-analytically.** Think carefully & systematically about whether & how data can improve business performance.

2. **Be able to interact competently on the topic of data mining for business intelligence.** Know the basics of data mining processes, algorithms, & systems well enough to interact with CTOs, expert data miners, and business analysts. Be able to envision data-mining opportunities.

3. **Have had hands-on experience mining data.** Be prepared to follow up on ideas or opportunities that present themselves, e.g., by performing pilot studies.
2. Instruction Method

This is primarily a lecture-based course, but student participation is an essential part of the learning process in the form of active technical and case discussion (please bring and use your name tags). The course will explain with real-world examples the uses and some technical details of various data mining techniques. The emphasis primarily is on understanding the application of data mining techniques, and secondarily on the variety of techniques and the mechanics of how they work.

Each class session has materials you must read prior to class. You should be prepared to be called on to discuss the readings.

Question Assignments

There will be a total of five question assignments, each comprising a (multi-part) question that will be given to you the week before the topic is discussed. Answers must be handed in prior to the class discussing the corresponding topic. They will be graded and returned promptly.

Case Studies

There will be two case analyses prepared by student teams. Students are encouraged to form a team of between 2 and 4 people. You can work as an individual (a team of 1 person) if you so desire, but teams tend to do a more comprehensive analysis than individuals. In data mining practice, teams with mixed business and technical expertise work best—if you can achieve that in your teams, all the better. You should decide on your teams by the end of the second class, and submit them to the instructor.

Teams are encouraged to interact with the instructor and TA electronically or face-to-face in developing their analyses. Half of the teams will present the first case in class, and half will present the second case. Selection of teams for presenting case #1 will be on a first-come-first-served basis (by email to the instructor), with (random) selection by the instructor if necessary.

Hands-on assignments

There will be four hands-on assignments—three individual and the fourth one in teams—based on data that we will provide. You will mine the data to get hands-on experience in formulating problems and using the various techniques discussed in class. You will use these data to build and evaluate predictive models. The fourth assignment will include a “competition”: one part of the data will be held back to evaluate the models you mine. The best-scoring team will get 5 extra-credit points.

For the hands-on assignments you will use the (award-winning) toolkit Weka.

http://www.cs.waikato.ac.nz/ml/weka/

A tutorial/demonstration on Weka will be given in class (at this writing class #3, but check the schedule to be sure). In order to use Weka you must have access to a computer on which you can install software (they don’t let us install software on the machines in the computer labs). If you do not have such a computer, please see the TA immediately so we can make alternative arrangements. The first hands-on assignment will be very easy, ensuring that you can install the software and get it running, before moving on to more challenging assignments.
4. Requirements and Grading
You should attend all class sessions, especially since the class meets infrequently and the sessions build on previous discussions.

You will hand in 5 brief answers to questions that will be assigned in class. Answers should be well thought out and communicated precisely. Points will be deducted for sloppy language and irrelevant discussion.

The two case studies require (i) analysis, and (ii) a critique of the analysis due in the class following the case discussion. The points to be addressed in the analysis will be discussed in class. The case studies will be handed out during the term. The analysis of the case is due (submitted to the instructor) one week prior to the case discussion. The analysis should be between 10 and 20 double-spaced pages. You must also submit a brief (1-2 page) self-critique of your case in the session following the case analysis.

Cases will be graded based on the initial write-up as well as the critique.

There will be no final exam.

The grade breakdown is as follows:
1. Questions (5): 25
2. Hands-on Assignments (4): 35
3. Case Studies (2): 30
4. Participation and Class Contribution: 10

Late Assignments

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.

5. Communication, Text, etc.

The Blackboard site for this course will contain lecture notes, reading materials, assignments, and late breaking news. It is accessible via: http://sternclasses.nyu.edu/

Post questions regarding course content to the Blackboard site (unless you are uncomfortable doing so) so that others can benefit from the answers. You are encouraged to contribute by answering/following up on posted questions.

Readings:

1. Textbook: available at the bookstore
   Data Mining Techniques, Second Edition
   by Michael Berry and Gordon Linoff
   Wiley, 2004
2. Supplemental readings will be provided as the class progresses.
3. The two cases (posted to the website).