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:


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. 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. You are expected to attend every class session, and to arrive prior to the starting time.

Homework Assignments
There will be a total of seven assignments, each comprising questions to be answered and some including hands-on tasks. The last assignment will be done in your teams (discussed below). Completed assignments must be handed in prior to the start of the class on the due date. If submitted by email, they must arrive at least one hour prior to the start of class. They will be graded and returned promptly.

The hands-on tasks will be 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 final assignment will include a “competition”: one part of the data will be held back to evaluate the models you mine.

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

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

Tutorials/demonstrations of Weka will be given in class (at this writing the first will be in 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 me 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.

Term Project
A term project report will be prepared by student teams. Student teams should be of 2 or 3 people. You should decide on your teams by the end of the third class, and submit them to me.

Teams are encouraged to interact with the instructor and TA electronically or face-to-face in developing their project reports. You will submit a pre-proposal for your project around mid-term. Each team will present its project in the last class meeting. We will discuss the project requirements and presentations in class.

Final Exam
The final exam date is noted in this document’s header and in the schedule of classes. The subject matter covered on the final will be discussed in class.
4. Requirements and Grading
You should attend all class sessions—the sessions build on previous discussions.

Answers to homework questions should be well thought out and communicated precisely. Points will be deducted for sloppy language and irrelevant discussion.

The points to be addressed in the term-project analysis will be discussed in class. The material needed for the term project will be handed out during the term. The analysis should be between 10 and 20 double-spaced pages.

The grade breakdown is as follows:
1. Homeworks (7): 35%
2. Term Project (1): 30%
3. Participation and Class Contribution: 10%
4. Final Exam: 25%

Late Assignments

Assignments are due prior to the start of the lecture on the due date. Turn in your assignment early if there is any uncertainty about your ability to turn it in on the due date. Assignments up to 24 hours late will have their grade reduced by 25%; 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. Material for the term project (posted to the website).

Please keep in mind the Stern Code of Conduct

http://w4.stern.nyu.edu/uc/currentstudents/codeofconduct.cfm?doc_id=5599