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

Businesses, governments, and individuals create massive collections of data as a by-product of their activity. Increasingly, decision-makers rely on intelligent technology to analyze data systematically to improve decision-making. In many cases, automating analytical and decision-making processes is necessary because of the volume of data and the speed with which new data are generated.

We will examine how data mining techniques can be used to improve decision-making. We will study the fundamental principles and techniques of data mining, and we will examine real world examples and cases to place data-mining techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science. In addition, we will work “hands-on” with data mining software.

The course is a combination of lecture, case studies, hands-on exercises and a final project.

Prerequisites: None

Course Objectives
After taking this course, you should:

1. Approach business problems data-analytically. Think carefully & systematically about whether & how data can improve business performance, to make better-informed decisions for management, marketing, investment, etc.
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, consultants, etc. Envision opportunities.
3. Have had hands-on experience mining data. Be prepared to follow up ideas or opportunities that present themselves, e.g., by performing pilot studies.

Focus and Interaction
The course will explain through lectures and real-world examples the fundamental principles, uses, and some technical details of data mining techniques. The emphasis primarily is on understanding the business application of data mining techniques, and secondarily on the variety of techniques. We will discuss the mechanics of how the methods work as is necessary to understand the fundamental concepts and business application.

I will expect you to be prepared for class discussions by having satisfied yourself that you understand what we have done in the prior classes. The assigned readings will cover the fundamental material. The class meetings will be a combination of lectures/discussions on the fundamental material, discussions of business applications of the idea and techniques, case discussions, and student exercises.

You are expected to attend every class session, to arrive prior to the starting time, to remain for the entire class, and to follow basic classroom etiquette, including having all electronic devices turned off and put away for the duration of the class (this is Stern policy, see below) and refraining from chatting or doing other work or reading during class. In general, we will follow Stern default policies unless I state otherwise. I will assume that you have read them and agree to abide by them

http://www.stern.nyu.edu/AcademicAffairs/Policies/GeneralPolicies/DefaultPoliciesforSternCourses/index.htm

Classroom equipment
Students will be required to bring their laptops during nearly all classes with wireless Internet access to complete the class demo and exercises.

Office hours and Email
Instructor’s office hours: Mondays 4-6 PM or by appointment.

If you have questions about class material that you do not want to ask in class, or that would take us well off topic, please talk to me after class or see me during the office hours.

You may also send emails to ask questions or set up appointments outside of office hours. Please type “NYU-DM” in the subject line of every email that you send.

I will check my email at least once a day during the week (M-F). Your email will get my priority when you put “DM3336” in the subject. If no reply is received within 48 hours, your email may have been overlooked and please feel free to send another email.

Course Homepage
The NYU Classes site is the main site for this Please check the page frequently for updates. You will find the following materials: syllabus, lecture schedule, lecture notes, group project, frequently asked
questions, and course resources. Please print out the necessary materials for yourself. I will assume that you have read all announcements and class discussion.

Course Materials and Textbook
Lecture slides and handouts distributed by the instructor. All lecture slides will be posted on NYU Classes website. You will be expected to flesh these out with your own note taking, and to ask questions about any material in the notes that is unclear after our class discussion. Depending on the direction our class discussion takes, we may not cover all material in the notes.


Grading
Student grades will approximately consist of the following elements as related to learning objectives

<table>
<thead>
<tr>
<th>Weekly assignments 20-40 points each</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final project presentation</td>
<td>20%</td>
</tr>
<tr>
<td>Final project report</td>
<td>20%</td>
</tr>
</tbody>
</table>

At NYU Stern we seek to teach challenging courses that allow students to demonstrate differential mastery of the subject matter. Assigning grades that reward excellent and reflect differences in performance is important to ensuring the integrity of our curriculum. Students generally become engaged with this course and do excellent or very good work, receiving As and Bs, and only one or two perform only adequately or below and receive Cs or lower. Note that the actual distribution for this course and your own grade will depend upon how well each of you actually performs this particular semester.

Weekly assignments
There will be homework assignments nearly every week. Each homework comprises questions to be answered and/or hands-on tasks. Except as explicitly noted otherwise, you are expected to complete your assignments on your own – without interacting with others.

Assignments will be due on Sunday night. Assignments will be graded and returned promptly. Answers to homework questions should be well thought out and communicated precisely, avoiding sloppy language, poor diagrams, and irrelevant discussion.
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.

For the hands-on assignments you will be using R and Rattle. R is a free open-source statistical analysis environment. You can download and install R from here: [http://www.r-project.org/](http://www.r-project.org/)

You will also want to install RStudio, a free development environment for R. [http://www.rstudio.com/](http://www.rstudio.com/)

Please plan to bring a laptop with R to every class session. If you do not have a laptop, please see me immediately so that we can make alternative arrangements.

**Late Assignments**

As stated above, assignments are to be submitted on NYU Classes on Sunday night. 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. Please turn in your assignment early if there is any uncertainty about your ability to turn it in on time.

**Final project**

Final project will be done by student teams. Team members will be assigned by the instructor. I will assign teams and distribute the outline for the final project by the third week of the semester. Teams are encouraged to interact with the instructor and TA electronically or face-to-face in developing their projects. You will submit a proposal for your project about half way through the course. Each team will present its project at the end of the semester. Each team will also submit a written report. We will discuss the project requirements and presentations in class.

**Regrading**

If you feel that a calculation, factual, or judgment error has been made in the grading of an assignment or exam, please write a formal memo to me describing the error, within one week after the class date on which that assignment was returned. Include documentation (e.g., pages in the book, a copy of class notes, etc.). I will make a decision and get back to you as soon as I can. Please remember that grading any assignments requires the grader to make many judgments as to how well you have answered the question. Inevitably, some of these go “in your favor” and possibly some go against. In fairness to all students, the entire assignment or exam will be regraded.

**Disability Policy**

If you have a qualified disability and will require academic accommodation during this course, please contact the Moses Center for Students with Disabilities (CSD, 998-4980) and provide me with a letter from them verifying your registration and outlining the accommodations they recommend. If you will
need to take an exam at the CSD, you must submit a completed Exam Accommodation Form to them at least one week prior to the scheduled exam time to be guaranteed accommodation.

**Honor Code**

We assume that you have complete integrity in all your class efforts. Violations of the University’s Honor Code will be taken extremely seriously, and they will be addressed promptly according to the established procedures. Students are to adhere to the Code of Student Conduct, and other policies and regulations as adopted and promulgated by appropriate University authorities. Copies of these documents may be obtained from the Office of the Dean of Students or from the offices of the academic deans. No infractions will be tolerated. Students violating the Code of Student Conduct will be dismissed from class and will receive an “F” for the course.

http://www.stern.nyu.edu/portal-partners/current-students/undergraduate/resources-policies/code-of-conduct

**Class Schedule (Tentative)**

<table>
<thead>
<tr>
<th>Week</th>
<th>Topics</th>
<th>Deliverables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction to Data Mining</td>
<td>CV</td>
</tr>
<tr>
<td></td>
<td>Exploratory analysis</td>
<td>Assignment 1</td>
</tr>
<tr>
<td>2</td>
<td>Data visualization</td>
<td>Assignment 2</td>
</tr>
<tr>
<td></td>
<td>Multiple linear regression</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Classification</td>
<td>Assignment 3</td>
</tr>
<tr>
<td></td>
<td>Logistic regression</td>
<td>Final project proposal</td>
</tr>
<tr>
<td>4</td>
<td>Dimension reduction</td>
<td>Assignment 4</td>
</tr>
<tr>
<td></td>
<td>Decision trees</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Model performance evaluation</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Neural networks</td>
<td>Assignment 5</td>
</tr>
<tr>
<td>7</td>
<td>SVM</td>
<td>Assignment 6</td>
</tr>
<tr>
<td>8</td>
<td>Ensemble techniques 1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Ensemble techniques 2</td>
<td>Assignment 7</td>
</tr>
<tr>
<td>10</td>
<td>Naïve Bayes</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Text mining</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Final group presentations</td>
<td>Final project presentation &amp; report</td>
</tr>
</tbody>
</table>