

Professor	<p>Claudia Perlich, PhD</p> <p>Adjunct Professor, Stern School of Business</p> <p>Email: claudia.perlich@gmail.com please put "NYU-DS" in the subject</p> <p>Phone: 914 4095609</p>
-----------	---

Course Description

The goal of this course is NOT to turn you into a data scientist or make you an expert on AI, LLM's or ChatGPT!, Rather, I will give you a deep enough understanding of the opportunities, techniques and critical challenges of using data science and machine learning in a business setting. It will be a hands-on experience, comparatively light on programming requirements. The focus is on the ability to understand and translate business challenges into data science problems and cover in depth the process of thinking about data sciences with respect to a specific problem that needs solving. A core requirement is the understanding of technical concepts and the personal experience in homework and projects of building data science solutions. We will examine how data science and predictive modeling technologies can be used to improve decision-making. We will study the fundamental principles and techniques of data science, and we will examine real-world examples and cases to place data-science techniques in context, to develop data-analytic thinking, and to illustrate that proper application is as much an art as it is a science. A large part of the class is the discussion of 'cases' as experiences by the teacher over her 20+ years of applying data science in a number of industries.

The course is a combination of lecture, case studies, homework and a final project.

Course Objective

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 science and machine learning for business intelligence.* Know the basics of data science processes, algorithms, & systems well enough to interact with CTOs, expert data miners, consultants, etc. Envision opportunities.
3. *Have had hands-on experience with machine learning.* Be prepared to follow up on 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 explore uses and some technical details of machine learning techniques. The emphasis primarily is on understanding the business application of data science techniques, and secondarily on the variety and technical details of specific algorithms. We will discuss the mechanics of how the methods work as is necessary to understand the fundamental concepts and business application. This is not an algorithm or a programming course! We will cover a number of cases during the course of the semester. Their goal is for you to understand how certain modeling decisions were made in the context of the business decision. Part of the homework's will be a review of cases and your understanding. I will expect you to participate in class discussions and be proactive about ensuring that you understand what we have done in the prior classes. The assigned readings will cover the fundamental material and I am always available to discuss.

Office Hours and Email/Phone

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 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). If no reply is received within 48 hours, your email may have been overlooked and please feel free to send another email. I strongly encourage you to see me at least once in this semester in my office hours. For urgent matters including setting up additional meeting times, you are welcome to call me or send me a text. Office hours will be held via zoom, the link will be in the calendar on Brightspace once the exact time and day are selected. The office hours are likely going to be during weekday evenings.

Course Coordination (Team, polls, office hours, etc.)

We will be using google docs, the link will be shared in a course announcement on Brightspace.

Course Homepage

The Brightspace site for this course will contain lecture notes, reading materials, assignments, extra-class discussions, and late-breaking news. You should check the site regularly. If you do miss one of the classes it is your responsibility to acquire the content. In particular, all classes are recorded and you can watch the last session. Important information and due dates will be communicated via announcements.

Lecture Notes and Readings

The textbook for the class will be:

Data Science for Business: Fundamental principles of data science and data analytic thinking
Provost & Fawcett (O'Reilly, 2013).

All slides will be posted on Brightspace shortly prior to the lecture. I may post some additional required readings as we go along. For those interested in going further, these following supplemental books give alternative perspectives on and additional details about the topics we cover – some of them are located in the Resources/Books folder. These are completely optional; you will not be required to know anything in these readings that are not in the primary materials or lectures. I have many other books that I can recommend, for example if you want a reference to a more mathematical treatment of the topics. Please don't hesitate to come and talk to me about what supplemental material might be best for you, if you want to go further.

- "Weka Book": [Data Science: Practical Machine Learning Tools and Techniques](#) by Ian Witten, Eibe Frank, Mark Hall
- [Elements of Statistical Learning](#) by Hasties, Tibshirany and Friedman ISBN-10: 0387848576
- [Introduction to Statistical Learning](#) with code examples in R
- [Data Science at the commandline](#) for those with interest in hardcore coding ...

Cases

This class does not cover cases in the classical sense but I will on post a business problem during class and expect you to talk about it. Data Science is a craft that requires doing in order to learn. Unfortunately, we have only limited time and I cannot expose you in depth to more than one project. Instead we will talk about these cases during class and I will share my experience in solving them. Some cases are full fledges publications and as the course progresses you should be able to understand nearly all of the technical details. Your participation in these discussions is crucial to your learning experience.

Software

During class we will use primarily WEKA and occasionally Excel (or another tool of your choice). WEKA is a GUI driven package with most major algorithms available. Instructions are on Brightspace.

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

Please download the version 3.8.5. During class we will have on occasion a "lab session" during which I will show how to do certain things in WEKA, so having it on the laptop that you are bringing to class will be useful. If you need additional help with using the data science software, please join me during office hours or plan on staying after class during one of the initial weeks. WEKA will be the tool you have to use for all homework assignments. For the project you are free to use ANY tool/software you wish. Recommended are Python, R, UNIX command line tools etc. WEKA is limited in its support of data pre-processing and formatting and is also unsuitable to deal with large datasets.

Homework Assignments

The homework assignments are listed in the class schedule below and are due as indicated on the class calendar. Final information will be communicated with announcements. Each homework comprises questions to be answered and/or hands-on tasks. Except as explicitly noted otherwise (see next paragraph), you are expected to complete your assignments on your own.

For the hands-on parts of the assignments (with Weka), I encourage you to initially work with your group members and other classmates to understand how to get Weka to do what you need to do, and then to complete your assignment on your own. So, for example, you could have a classmate help you do something similar, such that then you would be able to complete the assignment. Completed assignments must be handed on Brightspace unless otherwise indicated. The hands-on tasks will be based on data that we will provide. You will analyze 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. Plan on running into issues and start your homework early to be able to ask me questions PRIOR to the due date. Homework are a core learning experience - I prefer you to ask questions and get the homework right.

Late Assignments

I resume the right to reduce the grade for late assignments unless you have been in direct communication with me PRIOR to the due date - up to 24 hours late will have their grade reduced by 25%; assignments up to 5 days late will have their grade reduced by 50%. After 5 days, late assignments will receive no credit as I will be posting the solutions.

Term Project

The most important learning experience is the project. Please consider this a huge opportunity, not only to get a good grade in this course, but to shape your career as many students have done. It is an opportunity learn how to work with a team on data science problems, have a great entry on your resume, to potentially solve a cool problem and make a difference, and have me 'consult' on your problem for free.

The project will be executed by student teams of 3-4 students. *We will help seeing the group and you should decide on your teams by the end of the second week, and submit them to me – see the deadlines in the syllabus.* Teams are encouraged to interact with the instructor and TA electronically or face-to-face in developing their project reports. Each team will present its project at the end of the semester. We will discuss the project requirements and presentations in class. Deadlines for project deliveries will be communicated in announcement and will be shown in the class calendar on Brightspace. The dates in the syllabus are preliminary and subject to change.

I strongly advocate that you start your project early – you will have to find data as well as formulate the problem. The most time consuming by far will be managing the data. You are free to use any tool at your disposal and it is highly recommended that you form diverse teams with at least one person who can program and one with a strong business background.

Quizzes

We will have a 1 hour midterm and a 1.5 hour final. The subject matter covered and the exact dates will be discussed in class. The quizzes will test your understanding of the subject matter and your ability to express the learned concepts in a concise way. You will be able to access all course material during the quiz. However, using any other external sources (websites, ChatGPT, etc.) are NOT permitted and usage will result in penalties.

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., a photocopy of class notes). I will make a decision and get back to you as soon as I can. Please remember that grading any assignment 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.

FOR STUDENTS WITH DISABILITIES: 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 Accommodations Form to them at least one week prior to the scheduled exam time to be guaranteed accommodation.

betray

Requirements and Grading

The grade breakdown is as follows:

Homework	30%
Term Project	30%
Quizzes	30%
Class Participation/Case Discussion	10%

Stern grading policy requires specific percentage in the different grade levels and the total number of points will be rescaled accordingly. Please understand that I am not at liberty to make exceptions and that your grade unless you can show evidence of factual error in the grading is final.

Preliminary Class Schedule

Class	Date	Topic	Readings/ Preparation	Deliverables (Preliminary)
1		Introduction Terminology		(Homework 0)
2		Trees, Evaluation Basics WEKA 101	Read Chapter 1&2&3	Homework 1 Install WEKA
3		Optimization methods Classification Evaluation	Read Chapter 4&5	Homework 2 Project groups
4		Evaluation in Depth Medical Applications	Read Chapter 7&8,11	Homework 3
5		Naïve Bayes & Language IBM: Banter	Read Chapter 9 & 10	Homework 4 Project Proposal
6		kNN & Recommender IBM: Wallet & Artist	Read Chapter 6	Homework 5
7		Feature selection & creation Leakage IBM: Netflix	(Midterm Quiz)	
8		Causality		
9		Unsupervised: Associations & Clustering Fraud/Feightliner	Chapter 12 & 6	Project Update
10		Deployment & Deep Learning Proposal Evaluation		
11		Managing, Hiring, Ethics	Chapter 13 & 14 (Final Quiz)	
12		Project Presentation		Project Slides

Stern Mandated Syllabus Sections

Academic Integrity

We take pride in our well-rounded education and approach our academics with honesty and integrity. Indeed, integrity is critical to all that we do here at NYU Stern. As members of our community, all students agree to abide by the [NYU Academic Integrity Policies](#) as well as the NYU Stern Student Code of Conduct, which includes a commitment to:

- Exercise integrity in all aspects of one's academic work including, but not limited to, the preparation and completion of exams, papers and all other course requirements by not engaging in any method or means that provides an unfair advantage.
- Clearly acknowledge the work and efforts of others when submitting written work as one's own. Ideas, data, direct quotations (which should be designated with quotation marks), paraphrasing, creative expression, or any other incorporation of the work of others should be fully referenced.
- Refrain from behaving in ways that knowingly support, assist, or in any way attempt to enable another person to engage in any violation of the Code of Conduct. Our support also includes reporting any observed violations of this Code of Conduct or other School and University policies that are deemed to adversely affect the NYU Stern community.

Stern Code of Conduct

The Stern Code of Conduct and Judiciary Process applies to all students enrolled in Stern courses.

- For graduate students, information can be found here:
<https://www.stern.nyu.edu/uc/codeofconduct>.

To help ensure the integrity of our learning community, prose assignments you submit to NYU Brightspace will be submitted to Turnitin. Turnitin will compare your submission to a database of prior submissions to Turnitin, current and archived Web pages, periodicals, journals, and publications. Additionally, your document will become part of the Turnitin database.

General Conduct & Behavior

Students are expected to maintain and abide by the highest standards of professional conduct and behavior. Please familiarize yourself with Stern's Policy in Regard to In-Class Behavior & Expectations for Graduate students.

(<https://www.stern.nyu.edu/portal-partners/registrar/policies-procedures/general-policies/code-conduct>) and the NYU Student Conduct Policy (<https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/university-student-conduct-policy.html>).

General Student Accessibility

If you will require academic accommodation of any kind during this course, you must notify me at the beginning of the course and provide a letter from the Moses Center for Student Accessibility ([212-998-4980](tel:212-998-4980), mosescsa@nyu.edu) verifying your registration and outlining the accommodations they recommend. If you will need to take an exam at the Moses Center for Student Accessibility, you must submit a completed Exam Accommodations Form to them at least one week prior to the scheduled exam time to be guaranteed accommodation. For more information, visit the CSA website: <https://www.nyu.edu/students/communities-and-groups/student-accessibility.html>

Student Wellness

Our aim is for students to be as successful academically as they can, and to help them overcome any impediments to that. Any student who may be struggling and believes this may affect their performance in this course is urged to contact the Moses Center for Student Accessibility (see also the Student Accessibility section of this syllabus) at 212-998-4980 to discuss academic accommodations. If mental health assistance is needed, call the NYU's 24/7 Wellness Exchange hotline 212-443-9999.

Furthermore, please approach me if you feel comfortable doing so. This will enable me to provide relevant resources or referrals. There are also drop in hours and appointments. Find out more at <http://www.nyu.edu/students/health-and-wellness/counseling-services.html>

Graduate students can also reach out to the Academic Advising team at academicaffairs@stern.nyu.edu if you would like to receive more information or further support.

Name Pronunciation and Pronouns

NYU Stern students now have the ability to include their pronouns and name pronunciation in Albert. I encourage you to share your name pronunciation and preferred pronouns this way. Please utilize this link for additional information: [Pronouns & Name Pronunciation](#)

Religious Observance and Other Absences

NYU's [Calendar Policy on Religious Holidays](#) states that members of any religious group may, without penalty, absent themselves from classes when required in compliance with their religious obligations. You must notify me in advance of religious holidays or observances that might coincide with exams, assignments, or class times to schedule mutually acceptable alternatives. Students may also contact religiousaccommodations@nyu.edu for assistance.

LAPTOPS, CELL PHONES & OTHER ELECTRONIC DEVICES

The use of electronic devices (e.g., tablets or laptops), for the purpose of note-taking only, is permitted.