

Pratt logo Amir Imani, Visiting Assistant Professor

Pratt Institute School of Information

144 W. 14th St., 6th Floor

New York, NY 10011-7301

Email: [aimani@pratt.edu](mailto:aimani@pratt.edu)

URL: <https://amirosimani.github.io>

# Machine Learning Course

INFO 656-01 - Fall 2020

## Course Description

Machine learning is a rapidly growing field that develops algorithms for tasks such as data classification and prediction. These algorithms are programmed to operate and adjust themselves independently of human intervention (i.e., to learn), allowing data work to occur quickly and at scale. Machine learning is a key technology behind the automation across many social areas today, often branded AI.

This course offers an introduction to machine learning as a practical tool that we can use, and as a technological field with social implications. We will learn about key concepts in machine learning; survey a few key machine learning techniques, such as supervised methods for machine learning (regression and classification), which attempt to map data onto desired outputs, and unsupervised methods (clustering and association), which attempt to find structure within data itself; use openly available tools to implement these techniques on text and image data; learn how to assess the effectiveness of different techniques on particular datasets; and discuss basic issues that confront all machine learning methods.

Readings, class discussions, and hands-on sessions will be complemented by guest lecturers (TBD) from machine learning practitioners. Students will be assessed via a final project developed throughout the course, in addition to the project proposal, presentation, class participation, and lab assignments.

This syllabus is a living document; expect it to evolve over the course of the semester. All changes will be communicated in class and the updated syllabus will be uploaded on LMS. Since this is a new course, your participation and input will be crucial in shaping it to your needs. Feel free to ask questions and give feedback or suggestions, in person or via email, as we move into the semester.

## Course Schedule

Week	Session	Lab	Assignment
1	Introduction – What is Machine Learning?	Getting started with Python, Git	-
2	Data, Bias, Communication	Working with data in Python, Visualization	#1 TBD
3	Intro to Supervised Learning - Linear Models	Supervised learning in Python with scikit-learn	-
4	Supervised Learning cont. - Trees, Forests, Ensembles	Supervised learning in Python with scikit-learn	#2 TBD
5	Unsupervised learning - Clustering, Dimensionality Reduction	Unsupervised learning in Python with scikit-learn	-
6	Project planning	Project Workshop	Project Proposal
7	NO CLASS - Midterm break	-	-
8	Model Pipelines: Evaluation, Tuning, Feature Processing	End to End ML pipeline	#3 TBD
9	Neural Networks	Neural Network with PyTorch [TBD], Sequential models	-
10	Neural Networks - cont.	Neural Network with PyTorch [TBD], Advanced models: Convolution, Recurrent	#4 TBD
11	ML Ecosystem - Cloud, MLOps, Production, Pre-Trained Models, APIs	Cloud infrastructure (AWS) [TBD], Working with APIs	-
12	Human-Centric AI, Fairness	TBD	-
13	Labor and Machine Learning	TBD	-
14	Be a Good Engineer! Software Abstraction, Reproducibility, Open Source	TBD	-

Week	Session	Lab	Assignment
15	Industry and Career	TBD	-
16	Presentations	-	-

## Textbooks and Readings Materials

Throughout the semester, we will survey diverse perspectives about machine learning as a socially situated technology. The assigned readings will be complemented by in-class discussions, typically at the beginning of the class.

Throughout the semester, we will use the following books frequently:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). Chapter 8 (303-332)
- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists."
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

All reading materials and course slides (if applicable) will be provided as hyperlinks or downloadable files through Canvas.

Students will need a Google/AWS account for certain lab sessions. I believe the Pratt email address can serve this purpose, giving you access to Google Drive and Colab. Details on how to sign up will be provided as needed.

Additional resources including technical tutorials, example projects and datasets, resources about critical discourse, and more are listed on this GitHub repo, and will be updated as necessary: [https://github.com/Amirosimani/Machine\\_Learning\\_Pratt](https://github.com/Amirosimani/Machine_Learning_Pratt)

## Readings and Discussions

A provocation will include a summary of key points in the readings, as well as questions / observations you would like to raise or make. The provocations will serve as starting points of the in-class discussion and some of them will scaffold towards the project proposal and final project.

Students who are not motivators for the week are expected to complete the readings before class, and contribute to the discussion in class and/or online by replying to the forum thread.

### Week 1: Introduction – What is Machine Learning?

- Noah Lorang (2016) [Data scientists mostly just do arithmetic and that's a good thing](#) t.
- David Donoho (2017) [50 Years of Data Science](#), Journal of Computational and Graphical Statistics, 26:4, 745-766.
- Meredith Broussard, Artificial Unintelligence, ch.2-3 (13-39)
- Gideon Lewis-Kraus (2016) ["The Great A.I. Awakening"](#)

### Week 2: Data, Bias, Communication

- Paul Ford (2015) ["What is Code?"](#)
- Lisa Gitelman (ed.) (2013) ["Raw Data" Is an Oxymoron](#)". introduction (1-14)
- Andrey Kurenkov (2020) [Lessons from the pulse model](#)
- Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10), pp 1 - 23.
- Karen Hao (2019) [This is how AI bias really happens—and why it's so hard to fix](#)
- O'Neil, C. (2016). [The Ethical Data Scientist](#)
- Chun, W. H. K. (2005). [On software, or the persistence of visual knowledge](#). grey room, 26-51.

**Supplemental Material:** \* Konold, C., Finzer, W., & Kreetong, K. (2017). [Modeling as a Core Component of Structuring Data](#). *Statistics Education Research Journal*, 16(2). \* [Wrongfully Accused by an Algorithm](#) \* Prabhu, V. U., & Birhane, A. (2020). [image datasets: A pyrrhic win for computer vision?](#) \* Galloway, A. (2015). ["The Computational Decision"](#) \* Wong, D. M. (2010). The Wall Street Journal guide to information graphics: The dos and don'ts of presenting data, facts, and figures. WW Norton.

### Week 3: Intro to Supervised Learning - Linear Models

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). Chapter 2.1 (15-28)
- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists.". Chapter 2.1, 2.2 (27-84)
- Stephanie Yee and Tony Chu, ["A visual introduction to machine learning"](#)
- Gene Kogan and Francis Tseng, ["Fundamentals, introduction to machine learning"](#)
- Daly, L. (2017) ["AI Literacy: The basics of machine learning"](#)
- Elish, M.C. & Hwang T., [An AI Pattern Language](#). Chapter 1 (1-13)

**Supplemental Material:** \* Khan Academy, ["Least-squares regression equations"](#) (MOOC module, playlist)

### Week 4: Supervised Learning cont. - Trees, Forests, Ensembles

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). Chapter 8 (303-332)
- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists.". Chapter 2.3 (31-121)
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press. Chapter 5.1 (96-105)
- Chris Deotte, ["Classifier Playground"](#)

**Supplemental Material:** \* Elish, M.C. & Hwang T., [An AI Pattern Language](#). Chapter 2 (16-33)

### Week 5: Unsupervised learning - Clustering, Dimensionality Reduction

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). Chapter 10 (373-412)

- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists.". Chapter 3 (133-210)
- Kearns, M., & Roth, A. (2019). The ethical algorithm: The science of socially aware algorithm design. Oxford University Press., chapter 1

**Supplemental Material:** \* Nicolas Papernot, N., & Goodfellow, I. (2018) [Privacy and ML, two unexpected allies](#) \* Ji, Z., Lipton, Z. C., & Elkan, C. (2014). [Differential privacy and machine learning: a survey and review](#). arXiv preprint arXiv:1412.7584.

## Week 6: Project planning

- Patrick Hebron, "[Rethinking Design Tools in the Age of Machine Learning](#)"
- Lovejoy, J. & Holbrook, J. (2017). [Human-Centered Machine Learning](#) • Kearns, M., & Roth, A. (2019). The ethical algorithm: The science of socially aware algorithm design. Oxford University Press., chapter 2

**Supplemental Material:** \* Example projects and resources: <https://www.are.na/achim-koh/ml-design-ish> (The linked list is a preliminary one and will be updated; also, the examples are meant primarily as inspirations, and do not indicate what the final project should look like)

## Week7: NO CLASS - Midterm break

## Week 8: Model Pipelines: Evaluation, Tuning, Feature Processing

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). Chapter 2.2 (28-42)
- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists.". Chapter 5 (253-304)
- O'neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books. ch.1 (1-31) & ch. 8 (141-160)

## Week 9: Neural networks

- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists.". Chapter 2.3.8 (106-121)
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press. Chapter 6, 7.8
- The Coding Train, "[10: Neural Networks - The Nature of Code](#)" (YouTube playlist)
- 3Blue1Brown, "[Neural Networks](#)" (YouTube playlist)
- Daniel Smilkov and Shan Carter, "[A Neural Network Playground](#)"

**Supplemental Material:** \* Cathy O'Neil. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Broadway Books. Conclusion (199-218)

## Week 10: Neural Networks - cont.

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press. Chapter 9, 10, 11
- Chris Deotte, "[Neural Network Playground](#)"
- Daniel Shiffman, "[The Nature of Code](#)", ch.10
- Andrew Ng. [Nuts and Bolts of Applying Deep Learning](#)

**Supplemental Material:** \* [Machine Learning with Differential Privacy in TF](#)

## Week 11: ML Ecosystem - Cloud, MLOps, Production, Pre-Trained Models, APIs

- Shan Carter and Michael Nielsen, "[Using Artificial Intelligence to Augment Human Intelligence](#)"
- [MLOps](#)
- Andrew Ng. [Transfer Learning](#)
- Podcast: [Data Science in the Cloud](#) (with Paige Bailey).

**Supplemental Material:** \* Podcast: [Jitendra Malik: Computer Vision](#)

## Week 12: Human-Centric AI, Fairness

- Alex Galloway, "[Are Algorithms Biased?](#)"
- West, S.M., Whittaker, M. and Crawford, K. (2019). [Discriminating Systems: Gender, Race and Power in AI](#). AI Now Institute
- Blaise Agüera y Arcas, Margaret Mitchell and Alexander Todorov, "[Physiognomy's New Clothes](#)"
- Shannon Mattern, "[The Ethics of Automating Design](#)"
- Roelof Pieters and Samim Winiger, "[Creative AI: On the Democratization & Escalation of Creativity](#)"
- [coding is for everyone - as long as you speak english](#)
- [is your algorithm racist?](#)

**Supplemental Material:** \* Julia Angwin et al., "[Machine Bias](#)" \* Diana ben-Aaron, "[Weizenbaum examines computers and society](#)" \* Patrick Hebron, "[Rethinking Design Tools in the Age of Machine Learning](#)" \* Podcast: [Peter Wang (part 1/2): CEO/founder Anaconda, Creator of PyData]h(<https://podcasts.apple.com/us/podcast/peter-wang-part-1-2-ceo-founder-anaconda-creator-of-pydata/id1478016790?i=1000459268650>)

## Week 13: Labor and Machine Learning

- Siddhartha Mukherjee, "[AI Versus MD](#)"
- [where will predictive text take us?](#)
- [How three French students used borrowed code to put the first AI portrait in Christie's](#)
- Kate Crawford and Vladan Joler, [Anatomy of an AI System](#)
- Cade Metz, "[A.I. Is Learning From Humans. Many Humans.](#)"
- [Why Are Good Jobs Disappearing if Robots Aren't Taking Them? Part I and II](#)
- [Algorithmic Management in the workplace](#)

**Supplemental Material:** \* Shannon Mattern, "[Maintenance and Care](#)" \* Rosenblat, A. (2018). Uberland: How algorithms are rewriting the rules of work. Univ of California Press. \* [Sepsis Watch in Practice](#). \* Astra Taylor, "[The Automation Charade](#)"

## Week 14: Be a Good Engineer! Software Abstraction, Reproducibility, Open Source

- Nick Seaver, "[Knowing Algorithms](#)"
- Greg Wilson, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, Tracy K. Teal (2017). [Good enough practices in scientific computing](#)
- Causey, Tray. [Software development skills for data scientists](#)
- [Why Is Open-Source So Important? Part One: Principles And Parity](#)
- [Production ML systems](#)

**Supplemental Material:** \* Justin Kitzes, Daniel Turek, Fatma Deniz (2018). [The Practice of Reproducible Research](#) \* Cooper, Z. (2020) [Getting started with contributing to open source](#) \* [Docker for Data Science](#)

## Week 15: Industry and Career

- Shana Lynch, "[Andrew Ng: Why AI Is the New Electricity](#)"
- [Market your data science like a product](#)
- Caitlin Hudon (2018). [Imposter syndrome in data science](#)
- [Doing data science at Twitter](#)
- Podcast: [Getting Your First Data Science Job](#) (with Chris Albon).

## Week 16: Presentations

### Revisions to the Syllabus

All changes will be communicated in class and the updated syllabus will be uploaded on Canvas. This version is current as of August 14, 2020.

---

## Lab assignments

The latter part of each class will be a lab session related to the topic of the week. Sometimes, the lab session will be accompanied by a lecture-style session before it; in other cases, we will move into the lab session right after discussions.

At the end of some lab sessions, I will give you take-home assignments (4 total). The assignments will scaffold towards the final project. For example, you will be asked to explain some machine learning terminology or write code that does a specific task.

You need to submit your coding assignments through our dedicated GitHub Classroom. Details on how to submit the homework will be communicated in class. The homework assignments are due by the end of the day Monday before the next class.

I may also ask you to write down the amount of time you spent working on the assignment. This amount of time does NOT affect gradings in any way; I am asking in order to gauge whether I am giving you too much work or whether you are having trouble with some of the course material.

---

## Project proposal

I will ask you to choose a topic that you would like to explore in your final project, and to submit a proposal by mid-semester.

On week 6, we will have an in-class activity where you will share your idea(s) and give peer feedback. Your 800- to 1200-word proposal is due by the end of the following, and should include: • A description of the data you intend to use • A description of the machine learning task you intend to perform • A tentative and brief survey of existing work on the topic • A discussion of the significance of your topic

We will discuss the proposal in further detail in the coming weeks.

---

## Final project + presentation

Your final project is to run an experiment that applies a machine learning technique (such as classification, regression, clustering, etc.) that we learned on a dataset of your choice. You can design the project as a complete piece on its own, or as a component of a larger project.

Projects are due before the final class in the form of a write-up detailing your work process; you also need to submit the resulting model / dataset and code used. We will dedicate our final class to presentations.

A detailed rubric for the project and presentation will be distributed separately.

---

## Course Goal

The goals of this course are to:

- Introduce students to key concepts and some common techniques in machine learning, as well as openly available tools
  - Help students to develop technical and critical thinking skills regarding machine learning
  - Enable students to conduct a machine learning experiment and communicate the result of their project
-

# Student Learning Outcomes

By the end of the course, students will be able to:

- Describe different machine learning methods, including their limitations
- Select an appropriate machine learning method for a given use case
- Implement machine learning algorithms and assess their performance
- Execute a machine learning experiment using openly available tools
- Support the design of their experiment by discussing both the technical aspect and the topic’s significance

# Assessment and Grading

Lab assignments	20%
Participation (discussions and peer feedback)	30%
Project proposal	10%
Final project	30%
Presentation	10%

Pratt’s grading scale:

A	Superior work	4.0	96-100
A-	Superior work	3.7	90-95
B+	Very good work	3.3	87-89
B	Very good work	3.0	83-86
B-	Very good work	2.7	80-82
C+	Marginally satisfactory	2.3	77-79
C	Marginally satisfactory	2.0	73-76
F	Failed:	0.0	0-72

# Portfolio

Work completed for this course may be included in your portfolio. For more information on each program’s portfolio requirements, please visit the program’s respective webpage:

- MS Library & Information Science: Portfolio - <http://bit.ly/prattmslisportfolio>
- MS Information Experience Design: Portfolio - <http://bit.ly/prattmsixdportfolio>
- MS Data Analytics and Visualization: Portfolio - <http://bit.ly/prattmsdavportfolio>
- MS Museums and Digital Culture: Portfolio - <http://bit.ly/prattmsmdcportfolio> You are encouraged to meet with your advisor about including projects in your portfolio.

# Policies

This Course’s Attendance Policy

Students are allowed two individual absences for any reason; no notice or documentation is required. If you miss a session, be sure that you complete the readings, consult your classmates or the professor about the class, and (as always) demonstrate your knowledge of previous topics in later sessions. Upon a third or fourth absence, your grade will be lowered by one mark for each absence. Five absences will result in failure of the course (F). Also, please note that a grade of C- is considered failing at the graduate level.

Students who encounter long-term health issues should provide documentation to the Office of Health and Counseling and discuss their options with the professor.

Absences	Penalty	Example
0-2	None	NA
3-4	One mark per absence	A → A- (3rd absence) → B+ (4th absence) A- → B+ (3rd absence) → B (4th absence)

Absences	Penalty	Example
5+	Failure of the course	NA

For more information on Pratt Institute's Attendance Policy, please visit <http://bit.ly/prattattendance>.

#### Academic Integrity Code

Academic integrity at Pratt means using your own and original ideas in creating academic work. It also means that if you use the ideas or influence of others in your work, you must acknowledge them. For more information on Pratt's Academic Integrity Standards, please visit <http://bit.ly/prattacademicintegrity>.

#### Students with Disabilities and Accessibility

Pratt Institute is committed to the full inclusion of all students. If you are a student with a disability and require accommodations, please contact the Learning/Access Center (L/AC) at [LAC@pratt.edu](mailto:LAC@pratt.edu) to schedule an appointment to discuss these accommodations. Students with disabilities who have already registered with the L/AC are encouraged to speak to the professor about accommodations they may need to produce an accessible learning environment.

Requests for accommodation should be made as far in advance as reasonably possible to allow sufficient time to make any necessary modifications to ensure the relevant classes, programs, or activities are readily accessible. The Learning/Access Center is available to Pratt students, confidentially, with additional resources and information to facilitate full access to all campus programs and activities and provide support related to any other disability-related matters.

For more information, please visit <http://www.pratt.edu/accessibility/>.

#### Human Rights, Equity, BERT, and Title IX

Pratt Institute seeks to provide an environment that is free of bias, discrimination, and harassment. If you have been the victim of harassment, discrimination, bias, or sexual misconduct, we encourage you to report this.

If you inform me (your professor) of an issue of harassment, discrimination or bias, or sexual misconduct I will keep the information as private as I can, but I am required to bring it to the attention of the institution's Title IX Coordinator. You can access Title IX services by emailing [titleix@pratt.edu](mailto:titleix@pratt.edu). You can also speak to someone confidentially by contacting our non-mandatory reporters: Health Services at 718-399-4542, Counseling Services 718-687-5356 or Campus Ministries 718-596-4840.

In cases of Bias, this information may go to our Bias Education & Response Taskforce (BERT). You can contact BERT by either reaching out directly via [bert@pratt.edu](mailto:bert@pratt.edu) or by contacting the BERT Co-Chair and Title IX Coordinator, Dr. Esmilda Abreu.

For more information, please refer to the Community Standards webpage: <http://bit.ly/prattcommunitystandards>.