THE UNIVERSITY OF CHICAGO
HARRIS SCHOOL OF PUBLIC POLICY

PPHA 30545: Machine Learning for Public Policy

Dr. Christopher Clapp
Dr. Rosie Li

Syllabus, Fall 2026

Class Meetings (Clapp):	
Section 01 - TTh 9:30am-10:50am	Keller 0023
Section 02 - TTh 11:00am-12:20pm	Keller 0023
Class Meetings (Li):	
Section 03 - TTh 02:00pm-03:20pm	Keller 0023
Section 04 - TTh 03:30pm-04:50pm	Keller 0023
Lab Sessions:	
Lab 01 - F 10:00am-11:50am	Keller 0023
Lab 02 - F 10:00am-11:50am	Keller 0021
Lab 03, 04 - F 10:00am-11:50am	Keller 1022
Professors:	
Chris Clapp (he/him)	Keller 3039
Rosie Li (she/her)	Keller 3051

TAs: Kyra Sadovi (Head TA; she/her)
TBD

Course Description

It's an exciting time to study machine learning and data science more generally! We live in a digital era where many of our decisions and actions are tracked. Information is being produced and recorded at a stifling pace. While this may not seem novel to those who were born and have grown up in the Information Age, the amount of data available to researchers and policymakers is much more than what existed even a decade ago. Coupled with cheap computing power and expanded data storage, recent developments across statistics, computer science, and data-driven social sciences allow us to use all this data in a myriad of interesting ways. But what questions will we seek to answer with this newly available *big data* and these newly developed *machine learning* tools?

While machine learning tools are already being used extensively in marketing, finance, and business, their application to public policy is in its infancy (despite the techniques being the same across disciplines). Early examples of questions with policy implications include: can we predict unavailable data we take for granted in the developed world from available information in a developing world context? Is it possible to improve the accuracy of judges' bail decisions that hinge on whether the accused will commit additional crimes? Or can we inform doctors about the trade-offs inherent in prescribing potentially addictive opioids to patients for short-term pain relief by predicting who is likely to develop an addiction in the long run?

In order to ask and inform questions like these, this class will introduce you to ways to detect patterns in data, then use what you have learned to predict important outcomes or describe the salient relationships among inputs. While this requires an understanding of how and why these tools work, we will emphasize the intuition and application of these techniques over their theoretical underpinnings. We will do so by exploring nascent, policy-relevant applications of these methods, but, ultimately, the full impact of how these machine learning techniques inform and influence policy has yet to be determined. That's up to you!

Learning Objectives: "What's My Incentive for Taking This Course?"

Specifically, the purpose of the course is to introduce you to a wide array of the fundamental methods in modern machine learning. Each week, we will learn about and discuss a different set of techniques and their applications to public policy during lecture sections. During lab sessions, you will gain experience with those techniques by coding their implementation in Python.

Along the way you can expect to:

- Apply machine learning techniques to carry out policy-relevant analyses.
- Understand how the machine learning approach, which focuses on prediction, differs from the approach to fundamental statistical and/or causal inference you learned in Harris' core statistics classes.
- Gain an appreciation of why the bias-variance trade-off makes prediction inherently difficult.
- Recognize the different ways "long" and "wide" big data allow us to improve our predictions.
- Work with (often messy) real-world data.
- Continue developing your coding skills in Python as you learn new tools.
- Visualize, interpret, and convey your findings to audiences of different levels of technical sophistication.

The overall course objective is for you to be able to use machine learning tools to inform better policy and make the world a better place, as well as to become an informed and critical consumer of policy recommendations based on machine learning techniques. Additionally, the course will allow you to market your newly gained machine learning knowledge and skills when applying for jobs.

Prerequisites

In addition to assuming knowledge of the material in the "Harris Core" classes (Statistics for Data Analysis I & II), the prerequisite is:

- PPHA 30537 Data and Programming for Public Policy Programming.

This course is the second installment of the three-quarter core sequence of the Specialization in Data Analytics (https://harris.uchicago.edu/academics/design-your-path/specializations/specialization-data-analytics) at Harris. Students at Harris and from other parts of the University may enroll without having taken previous course(s) in the sequence *after students who have completed the prerequisite class have had a chance to enroll*. However, it is necessary for MPP students to take the full sequence in order to meet the necessary requirements for the specialization.

Considering the Course without the Prerequisites? For anyone who has not taken the prerequisites and is considering taking this course, first, thanks for your interest in our class! This course introduces machine learning techniques, then has students practice and apply them via Python coding-based labs, problem sets, and mini-projects. Based on this overview, there are two things you should consider. First, all machine learning is based on statistics, so it is in your best interest to have completed the "Stats Core" before taking this class. Second, while the class doesn't directly build on the material covered in the Data and Programming class (which teaches general coding skills in Python), you will be responsible for knowledge of the material covered in the prerequisite class(es). It is in your best interest to take this course without that prerequisite only if you have sufficient experience coding in Python *and* are aware that you may be at a bit of a disadvantage relative to the majority of the students in the class who are taking the Data Analytics sequence.

If you are considering taking the class without having taken the prerequisites, we recommend looking over the syllabi for those classes and making sure that you're comfortable with the topics and techniques that are covered before making your decision on whether or not to enroll. It may also be helpful to take a look at the textbook, which is available online (for free; see the "Materials" section of this syllabus for details on how to access it).

Evaluation

Your final grade in this course will be related to performance in several areas. The weight placed on each component will be as follows:

Attendance	10%
Problem Sets (4) & Mini-Projects (4)	40%
Assignment Quizzes	50%
Participation (Extra Credit)	02%

Attendance As per Harris policy (https://harris.uchicago.edu/student-life/dean-of-students-office/policies), regular in-person class attendance is required of all Harris students. It is also necessary (but not sufficient) to do well in the class. We will take attendance in each class beyond the first week, and only on-time, in-person attendance for the full class period counts towards your attendance grade. Please be sure to display your name tent in every class to facilitate the taking of attendance. Your attendance grade will be calculated as the percent of the total classes that you attend, but you are allowed two "free" absences before your grade is affected.[1] Per the Harris policy, we will notify the Dean of Students about students who are chronically absent.

Problem Sets and Mini-Projects There are four problem sets and four mini-projects in this class. Problem sets will consist of more structured questions (primarily) from the textbook. They are designed to help students cement their understanding of the conceptual material covered in lectures and get practice both applying the tools we learn and with coding. Mini-projects are designed to apply the machine learning concepts and tools covered in class to policy-relevant questions. As such, they are less structured, based on real-world data, and emphasize application to public policy over statistical concepts.

The focus of the class is on applying machine learning techniques, and the problem sets and mini-projects are designed with that goal in mind. These assignments will be graded for completion (as opposed to strict correctness of solutions), but you are expected to put forth a good faith effort on all questions. Remember,

[1] Please note that you do not need to notify us of short-term absences, and we cannot adjust the attendance policy on a case-by-case basis by excusing absences. We understand that students sometimes have legitimate reasons for being unable to attend class (e.g., illness), so you can miss class twice before it affects your grade. Only long-term issues of sufficient magnitude that warrant involving the Academic and Student Affairs team in the discussion can qualify for an exception to this policy.

these assignments are your chance to deepen your understanding of the techniques you'll be exposed to in class and demonstrate your ability to apply them. Both are valuable skills in the labor market and beyond!

You are welcome (and encouraged!) to form study groups (we recommend groups of no more than three students) to work on the problem sets and mini-projects together. But you must write your own code and your own solutions. Please be sure to include the names of those in your group on your submission. Please also be sure to practice the good coding practices you learned in the Data and Programming class and comment your code, cite any sources you consult, etc.[2]

All assignments will be submitted on Canvas via Gradescope. You may submit assignments late for up to 24 hours after the due date, but there will be a grade penalty for doing so. No assignments will be accepted after the 24 hour late period for any reason, valid or otherwise.[3] Not turning in an assignment, handing it in more than 24 hours late, or failing to turn it in before the Gradescope link expires will result in a grade of zero. We understand that students sometimes have legitimate reasons for being unable to complete assignments on time or give their full effort, so we will drop the lowest grade among these assignments when calculating your grade. Dropping the lowest assignment grade is intended to cover ordinary illness and other emergencies. Only long-term issues of sufficient magnitude that warrant involving the Academic and Student Affairs team in the discussion can qualify for an exception to this policy.

Assignment Quizzes

We will have short quizzes during lab sessions covering material from the previously submitted problem set or mini-project.[4] These in-person quizzes will be designed to assess your understanding of the material from the take-home assignments independent of help from classmates or any Artificial Intelligence (AI) tools. We have observed that students are often overconfident about their level of understanding when using AI tools relative to working without them, so you should optimize your approach to the assignments and preparation for the quizzes accordingly.

There will be no makeup quizzes. Again, only issues of sufficient magnitude that warrant involving the Academic and Student Affairs team in the discussion can qualify for an exception to this policy, but we will drop the lowest grade among these assignments when calculating your grade.

Class Participation Class participation points will be based on your level of active, attentive, inquisitive participation during in-class discussions and/or on the discussion board. For in-class participation, note that regular class attendance is generally a necessary (but not sufficient) component of earning in-class participation points. Additionally, to earn credit, you must record each instance of your participation (e.g., when you ask a question, provide an answer, contribute to a class discussion, etc.) using the submission form linked on the main Canvas course page.[5] Please submit a separate entry each time you participate. You only need a brief description of your question/answer/etc. (enough to jog our memory) and you should record all participation within 24 hours after class ends. You do not need to record participation via the discussion board - just your in-class participation!

[2] Part of demonstrating your ability to apply machine learning techniques requires using good coding style (in part because it makes it easier for the graders/potential employers to see that you understand what you're doing). So while coding style is secondary to applying the techniques you will learn, you should continue to use the good style you learned in the prerequisite class as much as is possible to make the code understandable. We may take points off if your code is hard to follow.

[3] Reasons include, but are not limited to: illnesses, athletic competitions, work trips, job fairs, job interviews, travel reservations, relative illnesses, relative funerals, out-of-town weddings, car accidents, car trouble, scooter trouble, tickets to see Billy Joel in concert, and emergency visits to the veterinarian with your dog.
This is both because we share answer keys after the late period and because students from advantaged social backgrounds are more likely to make requests for extensions, so being responsive to these requests can increase educational inequality.

[4] Note that the eighth and final quiz will take place during finals week.

[5] You will have to be logged into your UChicago Google account to submit a response.

We will supplement in-class participation with the Ed Discussion discussion board on Canvas. Please use the discussion board to post questions, discuss the material covered in the lectures or on the assignments, and answer questions posed by your peers. As being a good colleague is both an important way to have social impact and is valued by employers, participation points can be earned by making posts that are helpful to your peers.[6] While this can take many forms, points will primarily be awarded for answering classmates' questions on the discussion board. In doing so, you *may not* explicitly share more than 10 lines of code, provide step-by-step solution algorithms (e.g., pseudo code) to entire questions, or direct solutions. You may clarify ambiguities in the assignments, discuss conceptual aspects of lectures or problems, show output and error messages, and provide general guidance on how to correct errors in understanding or code.[7] Additionally, you may post brief summaries of news articles that describe applications of machine learning techniques to public policy relevant issues.[8]

Grades

This class requires a 60% or above to pass and is not curved. All passing letter grades will be determined based on the following intervals used in the Data Science Specialization sequence:

$$\text{A} \quad [95\% - 102\%] \mid \text{A-} \quad [90\% - 95\%) \mid \text{B+} \quad [85\% - 90\%) \mid \text{B} \quad [80\% - 85\%) \mid \text{B-} \quad [60\% - 80\%)$$

Pass/Fail (P/F), Withdrawal, and Incomplete grade requests will be handled in accordance with University and Harris policy. Students who wish to take the course pass/fail rather than for a letter grade must use the Harris P/F request form (https://harris.uchicago.edu/form/pass-fail) and must meet the Harris deadline, which is generally 9:00am on the Monday of the 5th week of courses. To earn a P grade, students taking the course P/F must: submit at least seven of the eight assignments, take at least seven of the eight quizzes, and earn a grade that is overall equivalent to at least a B- letter grade. Note that P/F grades or those below a B- do not meet the requirements for the Specialization in Data Analytics.

Materials

Data Analysis and Statistical Software

We will use Python software in this class, and you are required to code all assignments in Python. Python is free and open-source software that can be installed on all operating systems. The text editor or integrated development environment (IDE) you use for writing code for assignments is up to you.

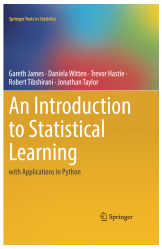
[6] Note that grades do not follow a curve in this class, so there is no penalty for helping others.

[7] For instance, a response to a peer that says, "to fix your error, the command should be '[...]'" is not permitted. Instead, saying, "I think you have a typo in the third argument of your command" is acceptable.

[8] Please note that in practice, the different means of class participation will be evaluated on an "either/or" basis. You are not required to participate in class via all possible modes of communication, although you are welcome to. There are multiple ways to participate because we want to give students as many opportunities to earn credit as possible, not because we want you to feel overwhelmed.

Textbooks

Required: *An Introduction to Statistical Learning with Applications in Python*, 1st Edition, by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. (ISBN-10: 3031387465)



You can download a free PDF of the textbook from the author's website: https://www.statlearning.com/. Note that there are two versions of the textbook. They differ only in the language used for the coding examples at the end of each chapter (Python or R). You will only need the Python-specific version for this class.

~~Office~~ Student Hours

Our (and the TAs') office hours for this class are listed on Canvas. The intent of office hours is part of the "hidden curriculum" for some students, so we want to clarify our expectations about how students should view and approach office hours. Those hours are for you, so please make use of them! You do not need to make an appointment to come to office hours; just drop by (be it with questions about course material, to discuss ideas, or just to chat). We will be available during those times.

Please make your best effort to attend during the posted times, but if you have a conflict or want to talk with one of us one-on-one, you are welcome to make an appointment for another time. We are happy to meet with students outside of office hours.[9]

Harris Tutoring Program

Harris offers free tutoring support for coding in Python, R, and Stata. See the Harris academic support programs (https://harris.uchicago.edu/student-life/dean-of-students-office/academic-support-programs) for more details. Any questions should be directed to your academic advisor or harrisdeanofstudents@uchicago.edu.

Course Policies

General

- Attendance is a requirement for Harris students and is part of how your grade will be determined. **That said, please do not attend class in person if you are sick!** We will record classes on Zoom so that students with short-term absences do not fall behind, but note that attending class remotely or watching recordings does not count towards your attendance grade.

[9] We only ask that you do your best to attend the regularly scheduled office hours since we have many students and there are economies of scale in the production of knowledge.

  - If you need a more-permanent remote learning accommodation, please contact the Dean of Students, Kate Biddle (kbiddle@uchicago.edu). Per Harris policy, all such requests can only be approved centrally, not by individual instructors.

- The use of electronic screens (laptops, phones) during class is prohibited, save for exceptions for students with a documented Student Disability Services (SDS) accommodation. This policy does not apply to hand-written note-taking on tablets that are laid flat on your desk. We reserve the right to modify this policy if we deem it necessary.

- The class webpage is available through the Canvas portal. We will use it to post announcements, assignments, and grades. Please check it regularly.

- Email, Canvas postings, and the discussion board are the official means of communication for out-of-class messaging. In other words, you are expected to check your UChicago email account and the Canvas site regularly.

- Email is inefficient. If you have a question about the class or the material, others probably do too! Questions and answers (knowledge) are public goods, so post your question to the discussion board, and feel free to answer questions your classmates ask. We will monitor and respond as well.

- If you have a question or concern about something you don't want to discuss publicly, feel free to email us. We will respond to email within 2 business days (Monday-Friday, 9:00am-5:00pm).

- Any and all results of in-class and/or out-of-class assignments and examinations are data sources that may be used in published research. All such use will always be anonymous, so the use of your data in any potential studies does not involve any risk to you beyond that of everyday life. The use of your data in a potential research study may not benefit you personally, but we may learn new things that could help others. By continuing in the class, you agree to allow your data to be used for potential future research projects. Participation is voluntary. To withdraw, simply email the professor indicating your desire to opt out. Withdrawing from potential research will involve no (class-related or other) penalty or loss of benefits to which you might otherwise be entitled.

Recording

- We will record all lectures. Recordings are intended as a complement to in-class instruction, not as a substitute for it.

- We will post recorded lectures only to Canvas in accordance with University and Family Educational Rights and Privacy Act (FERPA) guidelines.

  - The University has developed specific policies and procedures regarding the use of video/audio recordings (https://teachingremotely.uchicago.edu/recording-policy/).
  - FERPA is a federal statute that, broadly speaking, guarantees privacy over certain aspects of your educational records. You can view the details of the policy on the registrar's website (https://registrar.uchicago.edu/records/ferpa/).

- If you record a class, discussion section, office hours, or meeting without permission, or if you share any of the recorded videos without permission, you may be violating eavesdropping laws, copyright laws, or the FERPA statute. So do not post or share any such videos outside of Canvas. This also applies to any manipulated video.

AI Policy

The goal of the assignments in this course is not just to demonstrate that you can write code and answer questions based on the output, but to help you develop an understanding of complex concepts and associated critical thinking skills. This only comes from grappling with the material (both alone and in discussion with peers). AI tools, including large language models (LLMs), can be a complement or a substitute to this process. Examples of complementary uses of AI to learning include getting help understanding a concept or an explanation of an error message. Examples of using these tools as a substitute for learning include asking Copilot to write the code for a question or ChatGPT provide an interpretation of your results.

- While you are allowed to use AI tools in ways that are complements to your learning, you are NOT allowed to use AI tools in this class in ways that are substitutes for learning.
- We acknowledge that it is difficult to enforce this policy directly, so the in-person quizzes are designed to penalize students who use AI tools as a substitute for learning.
- To further encourage the appropriate use of AI, any time you get help from an AI tool, you must provide the query string you used and an explanation of how you used the AI tool's response as part of your answer.
- If you are unclear if a use is appropriate or if something is an AI tool, please check with us.

Academic Integrity[10]

We take the Harris Academic Integrity Policy (https://harris.uchicago.edu/student-life/dean-of-students-office/policies) very seriously. All students are expected to uphold the highest standards of academic integrity and honesty and comply with that policy. All students suspected of academic dishonesty will be reported to the Harris Dean of Students for disciplinary action. The disciplinary process can result in sanctions up to and including failing the class, suspension, or expulsion from the University. In addition, if in our judgment, the preponderance of the evidence indicates that a student has committed an honor violation on an assignment, at a minimum that student will receive an immediate grade of zero for that assignment and cannot earn a grade higher than a B- in the course, regardless of their performance on other assignments. This is regardless of the outcome of the disciplinary process. We trust every student in this course to fully comply with all of the provisions of UChicago and Harris' integrity policies. Here are specific expectations:

- On assignments and quizzes, it is expected that you will neither receive aid, give aid, nor access any materials other than what is explicitly outlined in the instructions.
- For problem sets and mini-projects, you may (and should!) work with other students, but it is expected that you will collaborate on all parts of the assignment (as opposed to the "divide and conquer" method).
- During the entire quarter, it is expected that you will not access old problem sets, mini-projects, answer keys, or any other class material at any time. This includes websites that post solutions under the guise of tutoring. (These sites both facilitate cheating and steal the intellectual property of the author.) This does not include the textbook authors' websites, Python documentation, or StackOverflow.
- During the entire quarter and thereafter, it is expected that you will neither post any class material on the internet nor share any class materials with other students through any other means. Furthermore, if you become aware that this has occurred, you are obligated to let us know immediately.

[10] We apologize for the heavy handed tone of this section. It is intended to protect the many honest students who take the class and academic integrity as a whole.

Americans With Disabilities Act

Students with disabilities needing an academic accommodation should contact UChicago's Student Disability Services (SDS). Please see their webpage for contact information (https://disabilities.uchicago.edu). If SDS determines a disability accommodation is appropriate, you should inform the Harris Dean of Students office by the end of the first week of class. The Harris Dean of Students office will work with the student and instructor to coordinate the implementation of the student's accommodations. Harris students are not required to submit their accommodations letter to the instructor, but please feel free to come talk to your instructor if you are comfortable doing so. We're happy to support your learning however we can.

Mental Health Services

Students differ in how much they know about mental health services. Your use of UChicago's Student Health and Counseling Services (SHCS) is free, confidential, and not linked to your academic file. If you find yourself suffering in silence, please do not hesitate to make use of the services provided by SHCS. Please see SHCS' mental health webpage for services and contact information (https://wellness.uchicago.edu/mental-health/). And if you are having serious mental, physical, or other problems, immediately contact the urgent medical care line at (773) 702-3625 (available 24 hours a day, 7 days a week).

Diversity and Inclusion

UChicago is committed to diversity and rigorous inquiry that arises from multiple perspectives, and Harris encourages thought-provoking discourse that involves not only speaking freely about all issues but also listening carefully and respectfully to the views of others. We concur with this commitment and view the diversity that students bring to our class as a valuable resource and a benefit to learning. We expect to maintain a productive learning environment based on open communication, mutual respect, and non-discrimination. We strive to present materials in a way that is respectful of diverse student backgrounds. As there can always be a gap between intent and execution, suggestions for promoting a positive and open environment are welcomed. Please feel free to correct us on your preferred name and gender pronouns if necessary.

Responsible Employees (Title IX)

All University of Chicago faculty and TAs are classified as "Responsible Employees." As such, they are required to report any discussions of sexual misconduct, dating violence, domestic violence or stalking to the Title IX Coordinator for the University. This includes the identities of the student making the complaint and alleged perpetrator. You will receive an email once a report is filed, but you are not obligated to meet with anyone or engage in the process. Alternatively, there are "Confidential Resource" employees at the University who do not have an obligation to share identifying information. For more information, including phone numbers, see the UChicago UMatter website (https://umatter.uchicago.edu/find-support/).

Syllabus Change Policy

Except for changes that substantially affect implementation of the evaluation (grading) statement, this syllabus is a guide for the course and is subject to change with advance notice.

Tentative Course Outline

The weekly coverage might change as it depends on the progress of the class. The "ISL" in the "Reading" column that follows is an abbreviation for the "An Introduction to Statistical Learning" textbook. "PS" is an abbreviation for "Problem Set," and "MP" is an abbreviation for "Mini-Project." Assignments will be due on Thursday evenings and quizzes will take place in lab sessions on Fridays (save for the final assignment/quiz during exam week).

Tentative Course Schedule

Week	Day	Topic	Reading	Due / Quiz
1	M/Tu	Introduction	ISL Ch. 1	
	W/Th	Statistical Learning	ISL Ch. 2	
2	M/Tu	Linear Regression & Moving Beyond Linearity	ISL Ch. 3 & 7	
	W/Th			
3	M/Tu			
	W/Th	Classification	ISL Ch. 4	PS 1
4	M/Tu			
	W/Th	Resampling Methods	ISL Ch. 5	MP 1
5	M/Tu			
	W/Th			PS 2
6	M/Tu	Linear Model Selection & Regularization	ISL Ch. 6	
	W/Th			MP 2
7	M/Tu			
	W/Th	Tree-Based Methods	ISL Ch. 8	PS 3
8	M/Tu			
	W/Th	Support Vector Machines	ISL Ch. 9	MP 3
9	M/Tu			
	W/Th	Deep Learning	ISL Ch. 10	PS 4
Exam Week				MP 4