Carnegie Mellon University
HeinzCollege

Machine Learning for Public Policy Lab 94-889
Fall 2024

Instructor: Mariana Escallon-Barrios
Email: mescallo@andrew.cmu.edu
Office: HBH 2110

Office Hours: Mondays 10:00-11:00 a.m. (HBH 2110)
Wednesdays 11:00 - 12:00 p.m. (HBH 2108)
By appointment via email

Teaching Assistants: Logan Crowl (lcrowl@andrew.cmu.edu)
Xueying Ding (xding2@andrew.cmu.edu)

Class Time & Location: Tuesday/Thursdays 11:00–12:20 pm (TBD)
Recitation Time & Location: Fridays 11:00–12:20 pm (TBD)

Course Description

This is a project-based course designed to provide training and experience in solving real-world problems using machine learning, with a focus on problems from public policy and social good.

Through lectures, discussions, readings, and project assignments, students will learn about and experience building end-to-end machine learning systems, starting from project definition and scoping, to modeling, to field validation and turning their analysis into action. Through the course, students will develop skills in problem formulation, working with messy data, communicating about machine learning with non-technical stakeholders, model interpretability, understanding and mitigating algorithmic bias and disparities, evaluating the impact of deployed models, and understanding the ethical implications of design choices made throughout the pipeline.

Pre-Requisites: Students will be expected to know Python (for data analysis and machine learning),SQL, and have prior graduate coursework in machine learning. This course assumes that you have taken graduate Machine Learning courses before and is focused on teaching how to use machine learning to solve real-world problems. Experience with git(hub) will be helpful and is highly recommended.

Course Resources

Textbook: The course will rely on selected readings from various sources and has no required textbook – each week, we'll have selected readings from a variety of sources, listed below.
Software: Students will be expected to store project code in a shared github repository, so you should create an account if you do not already have one (github.com). In the recitation session of week 2 the TAs will do a presentation on Github.

You may find a number of books useful as general background reading, but these are by no means required texts for the course:

- Data Science for Business by Provost and Fawcett
- Big Data and Social Science edited by Foster, Ghani, et al. Available online at https://textbook.coleridgeinitiative.org
- Practical Fairness: Achieving Fair and Secure Data Models by Nielsen
- Fairness and Machine Learning by Barocas, Hardt, and Narayana
- Weapons of Math Destruction by O'Neil
- Exploratory Data Analysis by Tukey

Additionally, the Global Communication Center (GCC) can provide assistance with the written or oral communication assignments in this class. The GCC is a free service, open to all students, and located in Hunt Library. You can learn more on the GCC website: https://www.cmu.edu/student-success/programs/communication-support/index.html.

Course Objectives

At the end of this course, you should be able to:

- Understand where machine learning fits as part of your analytical toolkit
- Understand the entire machine learning process (and get hands-on experience doing most of it)
- Learn how to use and evaluate machine learning methods (that you have covered in earlier classes) in the context of real problems.
- Be critical about the ethical implications of the use of machine learning in real-world applications.

Course Organization

In general, the course will be structured around three sessions each week:

During the Tuesday and Thursday sessions, we'll focus on structured lectures, discussions of the weekly topic, and students' presentations. As we move into the semester and the projects are underway, Thursday sessions will be reserved for group meetings and project work. Sometimes groups have trouble coordinating regular meeting times, so we want to find a way to dedicate some class time to help resolve this challenge.

During the Friday lab/recitation sessions, we'll discuss technical skills and tools you'll need for the project work early in the semester and then shift to check-ins with each team to discuss the status of their project work, generally surrounding short update assignments due on Wednesdays.

Although we're dedicating some time in class to work with your group, please note that successfully completing the project will require work outside of class time as well and will constitute the majority of the "homework" for the course.

Class expectations

Attendance: Everyone is expected to attend – and participate in – lectures. Students may miss **two** lectures with no penalty. If you anticipate more absences, please let me know as early as possible. Because this course is focused on discussion with your classmates, attending each session is important to both your ability to learn from the course and to contribute to what others get out of it as well.

Classroom behavior: Please arrive on time. If you arrive after the class has started, please enter quietly to avoid distracting other students and interrupting the lecture. If you arrive more than 20 minutes late to the class, it will count as a missed lecture. Laptops and iPads are permitted in class for taking notes or for class activities. Please refrain from using cell phones or electronics for other purposes during class. This might distract other students and keep you from fully understanding the material.

Participation: Participation is key to keep the class interactive. All questions are good questions. You are all here to learn and your question might help other students. We are all responsible for creating a safe and collaborative learning environment. For doing so, I expect you to respect each other ideas and learning process. All of you have different experiences and knowledge to bring to the class.

Communication: If you need to communicate with me, please come to my office hours or send me an email. If you have time conflicts with my office hours, email me and we can arrange another time. I am committed to replying to your emails within 24 hours on weekdays. For emails sent after 6:00 pm on Fridays, please allow until Monday morning.

Assessment

Project (50%): Throughout the semester, students will work together in small groups on an applied machine learning project that will illustrate the concepts discussed in class and readings. Please look at the project description for more details on the assignments and grade distribution.

Paper presentation and report (30%): Each student will have to present a paper and write a report of the presentation. In this applied class, we will cover multiple examples and these presentations will allow you and your peers to have a broader understanding of machine learning applications for public policy. Students will have one week to prepare for the presentation. The report should be submitted no later than 48 hours after the presentation. On Canvas you will find the details of the presentation.

Participation (10%): Participation will be graded based on participation in class activities and discussions. To make the class interactive, I expect all of you to come prepared to class, engage in class discussions, listen to your classmates, and share your thoughts from the readings and the class activities.

Attendance (10%): Students may miss two classes with no penalty. If you anticipate more absences, please speak with me as early as possible. For every missed class, excluding the first missed class, students will get a penalization in class attendance.

Regrade policy: We are committed to returning work to you as quickly as possible. Our goal is to grade assignments and return them within 1 week of the due date. All appeals for grading must be made within 2 weeks of the due date of the assignments. The appeal must be written and include the specific reason why you believe the grade is incorrect. You must submit the written appeal to me during office hours or after class.

Performance evaluation

The final grade for the course will consist of grades from the items mentioned above. Your final grade will be determined by your point placement on the following scale:

Graduate students:

94% or more	A	84%-86.9%	B	74%-76.9%	C
90%-93.9%	A-	80%-83.9%	B-	70%-73.9%	C-
87% -89.9%	B+	77%-79.9%	C+	0-69.9%	R

Undergraduate students:

94% or more	A	74%-83.9%	C	0-69.9%	R
84%-93.9%	B	70%-73.9%	D		

Be aware that I do not round up. That means that to pass the course you need 70% or more. This applies to all the items on the placement scale. For example, if you get a 93.99% in the class, your grade will be an A-. You need 94% or more to get an A.

Academic Integrity

In this class, you may use generative AI programs **ONLY** as a **tool** for your coding assignments. These programs can be powerful **tools** for learning and other productive pursuits, including the completion of some assignments in less time, helping you generate new ideas, or serving as a personalized learning tool.
However, your responsibilities as a student remain the same. You must follow the academic integrity guidelines of the university and of this class. If you use one of these generative AI tools to develop content for an assignment, you are required to cite the tool's contribution to your work. In practice, cutting and pasting content from any source without citation is plagiarism. Likewise, paraphrasing content from a generative AI without citation is plagiarism. Similarly, using any generative AI tool without appropriate acknowledgement will be treated as plagiarism. The university's policy on plagiarism applies to all uncited or improperly cited use of work, whether that work is created by human beings alone or in collaboration with a generative AI.

It is important that you recognize that large language models frequently provide users with incorrect information, create professional-looking citations that are not real, generate contradictory statements, incorporate copyrighted material without appropriate attribution, and can sometimes integrate biased concepts. Code generation models may produce inaccurate outputs. Image generation models may create misleading or offensive content.

While you may use these tools in the coding work you create for this class, it is important to note that you understand you are ultimately responsible for the content that you submit. Work that is inaccurate, biased, unethical, offensive, plagiarized, or incorrect will be penalized.

You may **not** use generative AI programs for the presentation, reports, or any other non-coding assignment. You are not allowed to use an AI generator to write or get ideas for any of your presentations or any of your written submissions. In accordance with CMU policy, *it is the responsibility of each student to produce her/his/their own original academic work.* (https://www.cmu.edu/policies/student-and-student-life/academic-integrity.html) If there is evidence that you copy from an AI generator, you will receive zero credit for the assignment and face academic sanctions according to the University policies. You are encouraged and expected to familiarize yourself with the University policy. If you have any questions please contact me or the TA.

Other information

Accommodation

If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.

Well being

School can be challenging at times. Make sure to take time for yourself and do activities you enjoy. The University Provost provides the following thoughts that I believe are important to share.

Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is almost always helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website athttp://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

Inclusion and Diversity

We must treat every individual with respect. We at Heinz College are diverse in many ways, and this diversity is fundamental to building and maintaining an equitable and inclusive campus community. Inclusion is ensuring everyone feels respected in the environments they interact with other. Diversity, can be broadly defined as the presence of people who identify with multiple attributes, these include but are not limited to race, color, national origin, language, sex, disability, age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information. Each of these diverse identities, along with many others not mentioned here, shape the perspectives our students, faculty, and staff bring to our campus. I will work to promote and create an inclusive and equitable learning environment. I understand that diverse backgrounds require different learning tools and I am committed to provide each of you with the tools you need to grow and learn in my class.

We, at CMU, will work to promote diversity, equity and inclusion not only because diversity fuels excellence and innovation, but because we want to pursue justice. We acknowledge our imperfections while we also fully commit to the work, inside and outside of our classrooms, of building and sustaining a campus community that increasingly embraces these core values. Each of us is responsible for creating a safer, more inclusive environment. The university encourages anyone who experiences or observes unfair or hostile treatment on the basis of identity to speak out for justice and support, within the moment of the incident or after the incident has passed. Anyone can share these experiences using the following resources:

- Center for Student Diversity and Inclusion: csdi@andrew.cmu.edu, (412) 268-2150
- Ethics Reporting Hotline. Students, faculty, and staff can anonymously file a report by calling 844-587-0793 or visiting cmu.ethicspoint.com.

All reports will be documented and deliberated to determine if there should be any following actions. Regardless of incident type, the university will use all shared experiences to transform our campus climate to be more equitable and just.

Other university resources

Student academic success center: The center supports student success through a number of group and individual programs. You can find more information about their programs at https://www.cmu.edu/student-success/index.html. If you are unsure of what programs will best suit your needs, you can sign up for a consultation with a staff member to identify the support you need.

CMU food pantry: Free resource committed to reducing hunger among students. For more information, visit https://www.cmu.edu/student-affairs/resources/cmu-pantry/ or contact cmu-pantry@andrew.cmu.edu

Emergency support funding: The university has several resources to provide emergency funds for students in times of need. More information about these resources and the application process is available at https://www.cmu.edu/student-affairs/dean/loans/

Course Plan

There may be adjustments to the schedule throughout the semester. Please check this document and Canvas regularly to keep up to date with the class schedule.

Week	Dates	Topics	Additional resources
1	08/27 08/29 08/30	Intro and Project Overview Scoping, Problem Definition, Balancing goals Meet your team	
2	09/03 09/05 09/06	Case studies and discussion Acquiring data, Privacy, Record Linkage GitHub tutorial	
3	09/10 09/12 09/13	Data exploration Analytical formulation of policy projects Project work	
4	09/17 09/19 09/20	Building ML pipelines Working on your project pipeline Project work	
5	09/24 09/26 09/27	Performance metrics Project work Project work	
6	10/01 10/03 10/04	Performance metrics part 2 Temporal validation deep dive Group check-in	
7	10/08 10/10 10/11	Feature engineering Project work Group check-in	
8	10/18	Fall Break - No class	
9	10/22 10/24 10/25	No class on Tuesday ML Modeling in Practice Group check-in	
10	10/29 10/31 11/01	Project presentations part 1 Project presentations part 2 Group check-in	
11	11/05 11/07 11/08	Democracy day - no class Performance metrics Group check-in	
12	11/12 11/14 11/15	Model interpretability Ethics workshop Group check-in	
13	11/19 11/21 11/22	Bias and fairness part 1 Bias and fairness part 2 Group check-in	
14	11/26 11/29	Class wrap up Thanksgiving break	
15	12/03 12/05 12/06	Project work Presentations Presentations	