Let’s do a quick poll

https://orie5355.github.io/Fall_2026/attendance/

Please use your **cornell netid – correct netid is important for attendance grading**

(for example, ng343).

ORIE 5355: Applied Data Science - Decision-making beyond Prediction

Lecture 1: Introduction

Nikhil Garg

Plan for today

- Content overview
- Syllabus/Course structure
- Questions
- Time permitting, move on to first content lecture
- Hopefully end a few minutes early for specific questions

Please interrupt with questions at anytime

(but raise your hand)

Who are we?

Instructor: Nikhil Garg

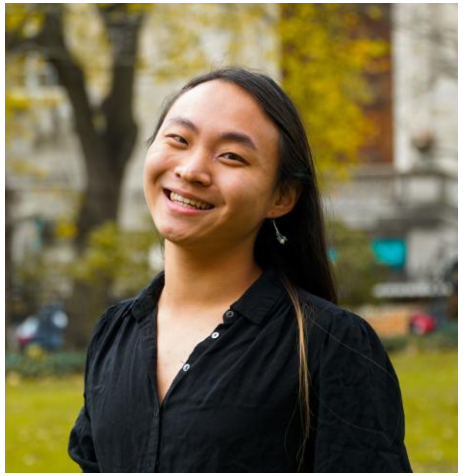
Asst Professor, Cornell Tech, ORIE

Research on the application of algorithms, data science, and mechanism design to the study of democracy, markets, & societal systems

Past experiences/collabs: Uber, Upwork, other marketplaces, campaign data science, many city agencies

TAs:

Evan Dong, PhD Student, CS



Vyoma Raman, PhD Student, CS

Content overview

What is this class *not* about

Intro to data science → **Urban Data/Data Science in the Wild**

How to train many types of cool prediction models → **Applied Machine Learning**

Deep learning/LLMs → many other courses

Theoretical analysis of markets (pricing, queuing) → **Service Systems and Online Markets; Networks and Markets**

What *is* this class about?

Let’s assume you know the machinery for training **ML** models

Then, how do you make + evaluate algorithmic **decisions**, in the presence of

- Data constraints – censoring, selection effects, limited data
- Business/resource constraints – capacities, communication limits
- Competition
- Changing environments
- (Perceived) fairness and ethical constraints
- User incentives, strategic behavior, adversarial behavior

Challenges *before* and *after* you train models

Who is this course for?

You know (or are learning right now) how to train and evaluate (basic) machine learning models

You want to learn how to actually use such these models in:

- Online platforms (Uber, Etsy, Netflix, Stripe…)
- Government (transportation, polling, census, providing services)

…any data science setting about people

Organization

Data (~2-3 weeks): What is it, how to collect it, and common challenges

Analysis (~5 weeks): Common tasks in algorithmic-societal systems

- Recommendation (~2 weeks)
- Pricing (~2 weeks)

Experimentation (~2-3 weeks): How do we know a model is good?

Miscellaneous (~2-3 weeks)

Data collection and processing

- What *is* data? Where does it come from? What does it *represent?*
- Common challenges in data collection

  Selection biases, censoring, and other challenges
- Polling/surveys as an extended example
  - What goes wrong in measuring opinions (mean estimation)
  - Some techniques that somewhat work
  - US 2016 election polls as a case study
- Other challenges and contexts: online ratings, privacy, etc.

Analysis: Recommendation

- Basics of recommendation: collaborative filtering, matrix factorization, and content-based recommendations with LLMs
- Individual- vs demographic-based personalized recommendations – tackling the cold start or low data regime in practice
- Other challenges for recommendation in practice:
  - Matching and capacity constraints
  - Two-sided fairness
  - How censored data + feedback loops affect recommendations
  - Incorporating human agency and expressed preferences

Analysis: Pricing

At what price do I sell my items?

- Optimal pricing given market data
- Personalized pricing – individual vs demographic based
- Dynamic pricing – pricing over time
- Exercise on when personalized and dynamic pricing is ethical
- Case studies:
    - prices and wages in online marketplaces
    - congestion pricing

Experimentation

How do I know if my product/intervention/action/treatment works?

- A/B testing basics: clinical trials, standard experimentation
- Why standard techniques fail in people-centric systems: interference
- Experimentation in practice
  - Dealing with networks and interference
  - Experimentation over time
  - Experiments in 2-sided marketplaces
  - Running many experiments
- Case studies: Covid/clinical trials, online marketplaces

Miscellaneous

[Exact topics TBA]

- Differential privacy: how do I share personal data in a “optimally” private manner?
- Deeper dive into fairness audits: how do I measure whether my algorithm is “unfair”?
- Algorithmic explainability and transparency
- “Human-in-the-loop” machine learning and Human-AI collaboration
- Steering LLM models and recommender systems

Class themes

- The right data and interpretation beats modeling sophistication
- The solution is not (always) technical
- You're not done after training a model, and you don't start there either
- Most mistakes are made in understanding and applying the basics
- Domain expertise is essential
- Design with privacy, ethics, and fairness in mind – not as an afterthought
- Historical performance on training set is often a terrible measure
- Concepts, not (just) methods
- Be curious!

Syllabus

https://orie5355.github.io/Fall_2026/syllabus/

Assignments + Grading

Homework: 30%. 3 homeworks. Each HW is an equal part of the homework grade. Lowest score replaced by quiz average.

- Both written questions and Python programming
- Largely a learning tool and graded primarily for completion + understanding

In class quizzes: 30%. 4 in class quizzes. Lowest score dropped.

Final project: 25%.

- Primarily programming

Participation: 15%.

- Including in-class polling/questions

Note the late day policy in the syllabus

Final project

- Most fun part of the class – a chance to apply everything we've learned
- You'll take the role of one entity in a marketplace competing to sell items against other students

Prediction, pricing, recommendations, game theory, experimentation

Attendance

- **Don't come to class sick or if you suspect you're sick**
- No remote participation
- Will often do in-class polling, contributes to participation grade
- Mandatory in-person attendance for guest lectures and a couple other days (unless sick, religious observance, or other SDS accommodation) – almost always take attendance
- TA Office hours will be partially remote, but will not be recorded

Do not fill in the attendance sheets unless you are here in person – substantial penalty in overall course grade if you do

Course communication

EdStem Discussion: First resource for any question

Office hours: You are strongly encouraged to come to office hours for any reason.

Email: Only for private questions and concerns. Technical questions will not be answered over email – please use Ed Discussion.

Classroom norms

- Take space, make space: allow others to join the conversation.
- Embrace a growth mindset. Mistakes are a valuable learning opportunity.
- Ask questions!
- Be willing to give and receive feedback respectfully.
- Recognize that we come from different disciplines and have different academic experiences. Be ready to explain concepts and terms.

Important links + resources

- [Course website]
- [Canvas] – Rarely used, except sharing private files
- [Ed Discussion] – Primary communication tool
- [Calendar] – Class google calendar with lectures, OHs
- [Gradescope] – Place to turn in all assignments
- [YouTube] – Lecture recordings from 2021.

Announcements

- This week: my office hours Wednesday after class
- Course pre-survey, posted on Ed discussion
- Office hours preference form posted online
- TA office hours start next week
- Homework 1 will be posted this week or next week

Questions?