ORIE 5355: People, Data, & Systems

Lecture 17: Prediction about people is hard

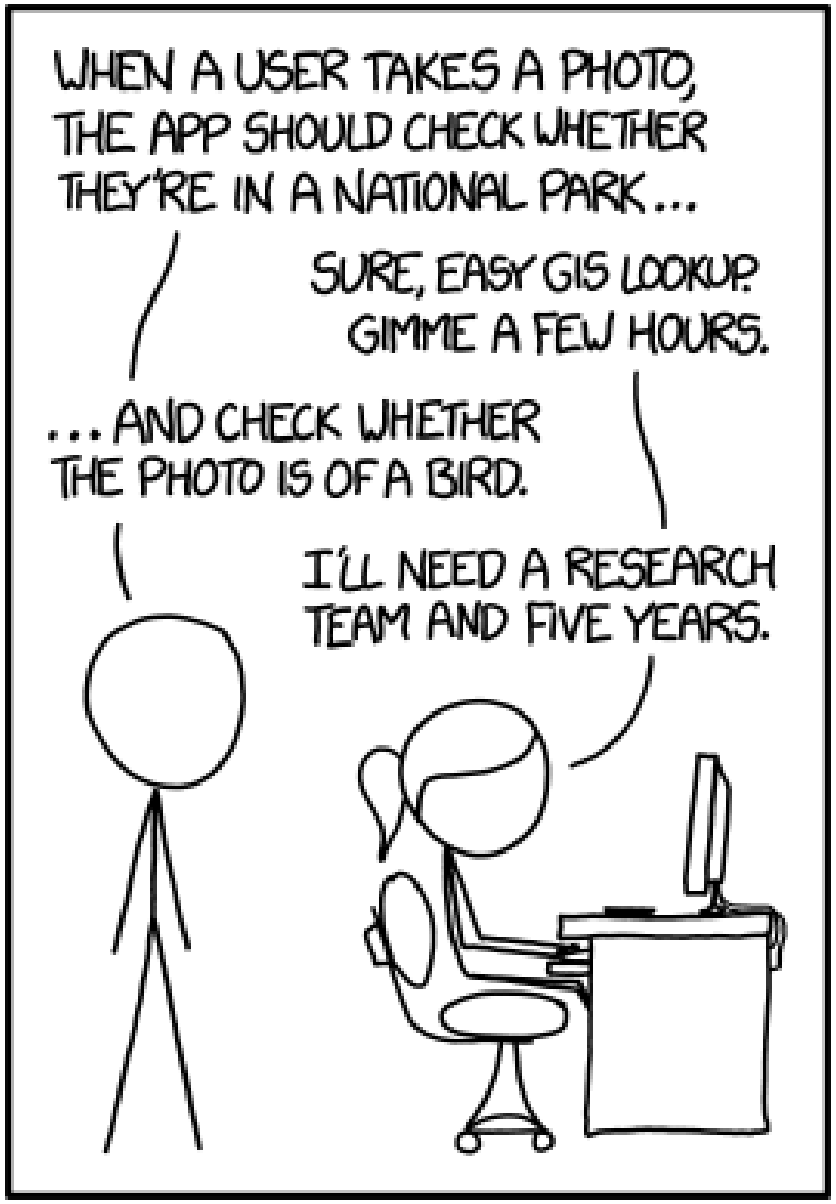
Nikhil Garg

Course webpage: https://orie5355.github.io/Fall_2021/

Announcements

- Part 1 of project due tomorrow night – NO LATE DAYS
- Regular office hours today (Zhi)
- No office hours Wednesday or Friday this week (Thanksgiving)
  - No class Wednesday
- Back to standard in-person lectures on Monday
- Details on turning in Project Part 2 posted soon
  - For now, make sure your code can run on Google Colaboratory with default packages
  - Cannot take more than ½ seconds per customer
  - Code + any needed data (e.g., a pickled trained model can be stored) in less than 10MB

“Prediction is hard, especially about the future”



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

“Measuring the predictability of life outcomes with a scientific mass collaboration”

Matthew J. Salganik, Ian Lundberg, Alex Kindel, Sara McLanahan, et al. 2020

Research question and method

Question: How predictable are life trajectories, if you have a *bunch* of data about someone's early childhood?

Method: "Common challenge" methodology – give a bunch of researchers the same data and same task. Then, see how they perform in predictions.

Why?

- Suppose only 1 research team. Maybe they were bad at machine learning?
- "File drawer" effect in research publications

Data set: ”Fragile Families & Child Wellbeing Study”

- ~5000 families in a large US city, each who gave birth to a child around the year 2000
- Collected when child was ages 1, 3, 5, 9, and 15
- Extremely rich dataset: interviews with parents, teachers, child themselves; home visits

  “The FFCWS consists of interviews with mothers, fathers, and/or primary caregivers … The parent interviews collected information on attitudes, relationships, parenting behavior, demographic characteristics, health (mental and physical), economic and employment status, neighborhood characteristics, and program participation. At ages nine and fifteen, children were interviewed directly (either during the home visit or on the telephone). The direct child interviews collected data on family relationships, home routines, schools, peers, and physical and mental health, as well as health behaviors.”

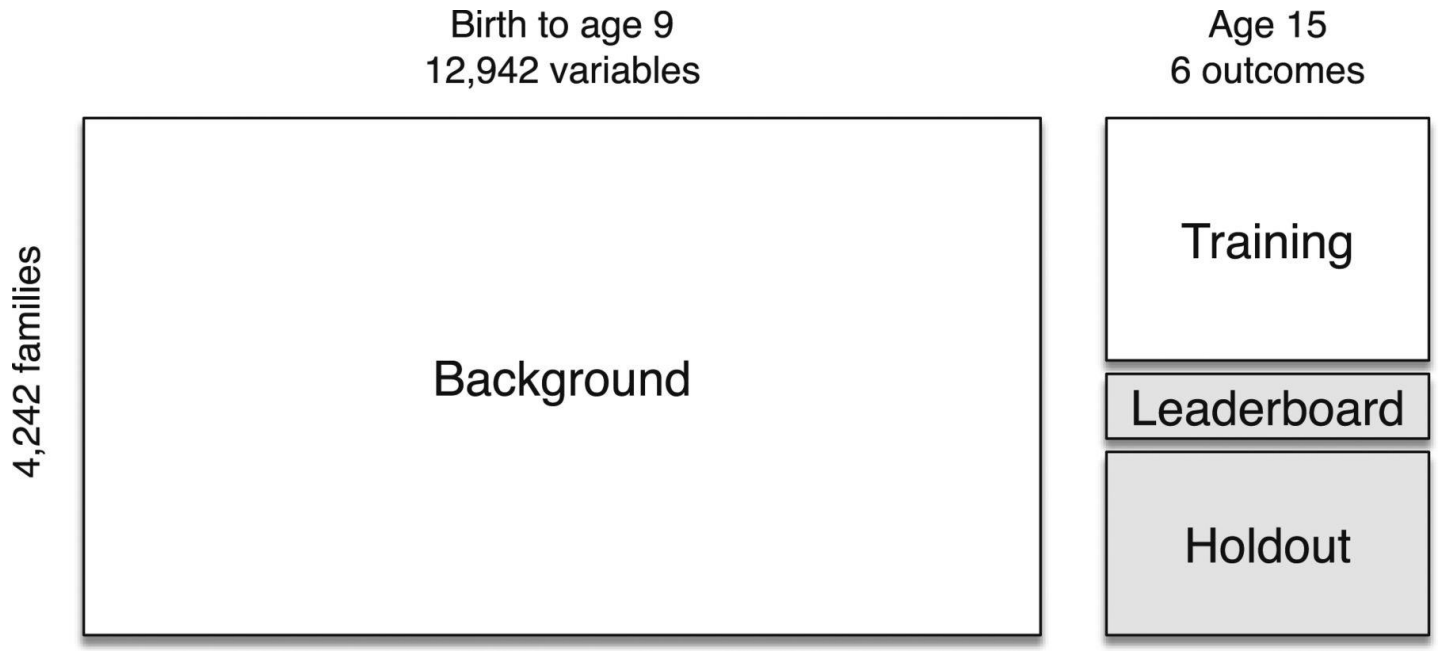
  The In-Home Study collected information on a variety of domains of the child's environment, including: the physical environment (quality of housing, nutrition and food security, health care, adequacy of clothing and supervision) and parenting (parental discipline, parental attachment, and cognitive stimulation). In addition, the In-Home Study also collected information on several important child outcomes, including anthropometrics, child behaviors, and cognitive ability. This information was collected through: interviews with the child's primary caregiver, and direct observation of the child's home environment and the child's interactions with his or her caregiver.

Machine learning task

- Try to predict 6 outcomes for child at age 15: “1) child grade point average (GPA), 2) child grit, 3) household eviction, 4) household material hardship, 5) primary caregiver layoff, and 6) primary caregiver participation in job training”
- Data given to researchers:
  - Background dataset: "4,242 families and 12,942 variables about each family,” from the Fragile Families dataset at ages 1, 3, 5, 9
  - Training dataset for the 6 outcomes (age 15) for half the families
- Not given/publicly available: age 15 outcomes for rest of families

Common challenge approach

- [Essentially a Kaggle competition/class project Part 1]
- Give the test dataset, ask researchers to predict outcomes for the test dataset; host a “leaderboard” for how good people are doing
- At some point, end the competition and see how everyone did



How good were [the best] predictions?

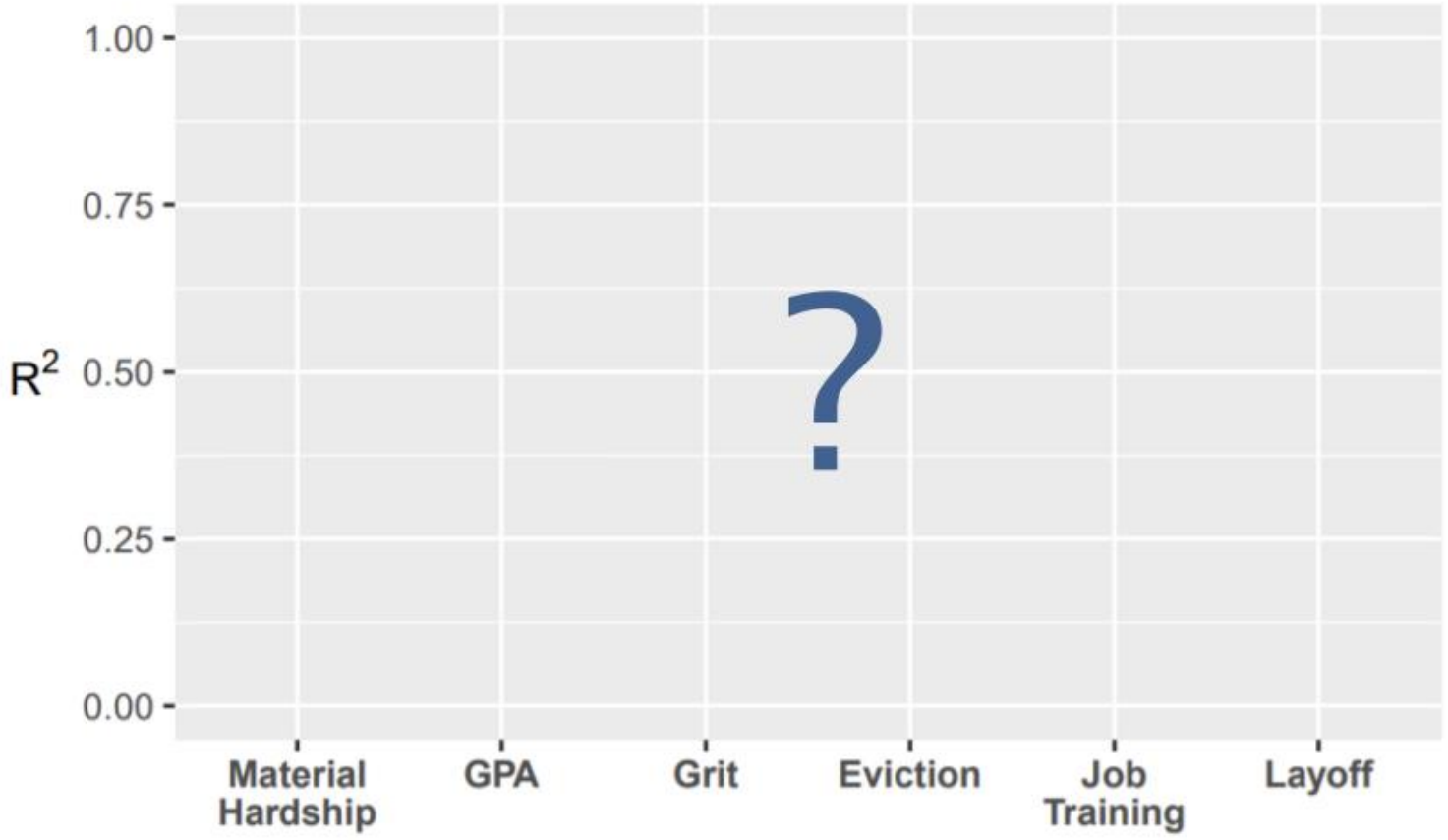


Image credit: Matthew Salganik, via Arvind Narayanan talk here: [MIT-STS-AI-snakeoil.pdf (princeton.edu)]

How good were [the best] predictions?

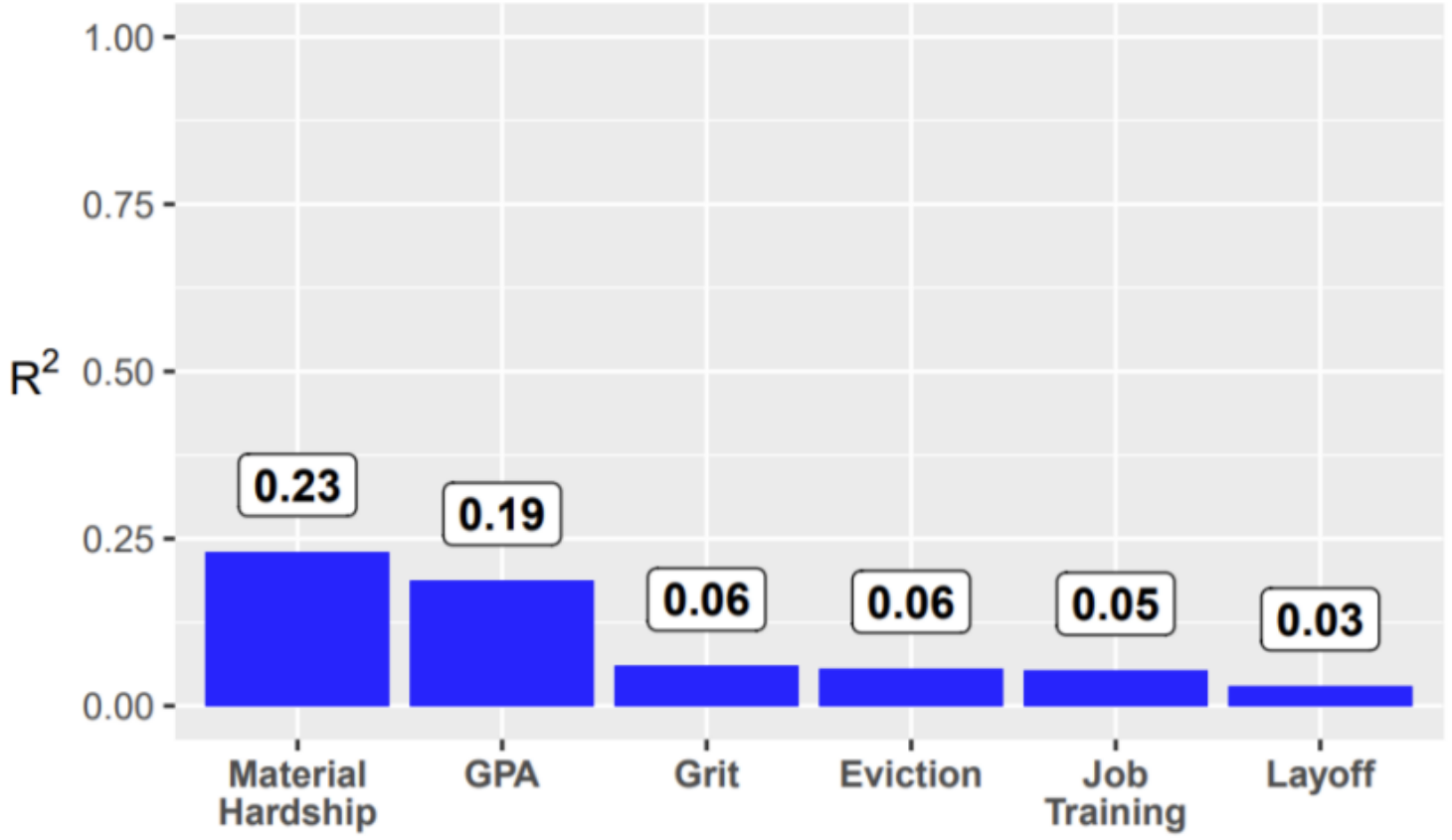


Image credit: Matthew Salganik, via Arvind Narayanan talk here: [MIT-STS-AI-snakeoil.pdf (princeton.edu)]

Linear regression as good as best predictors

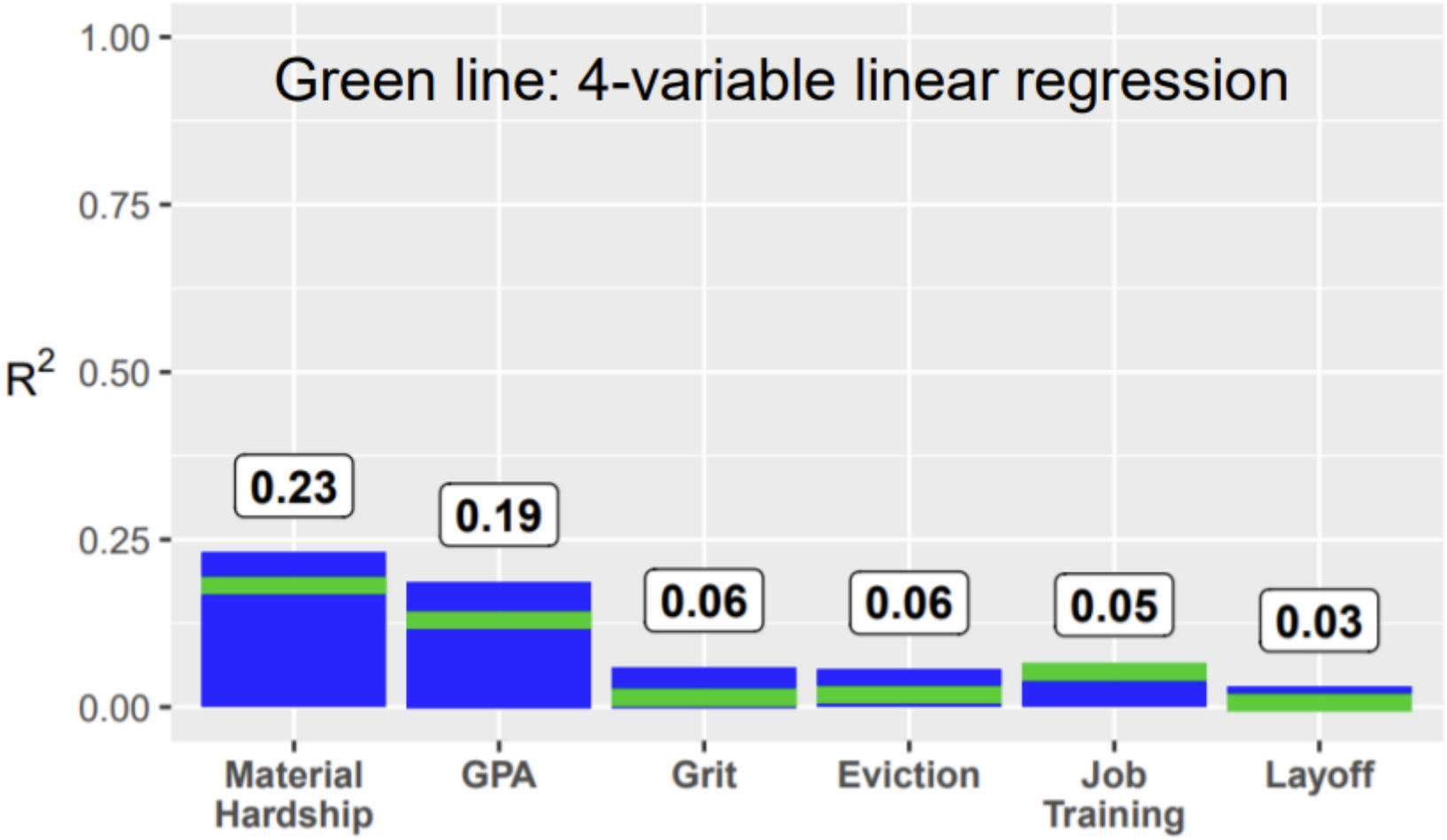
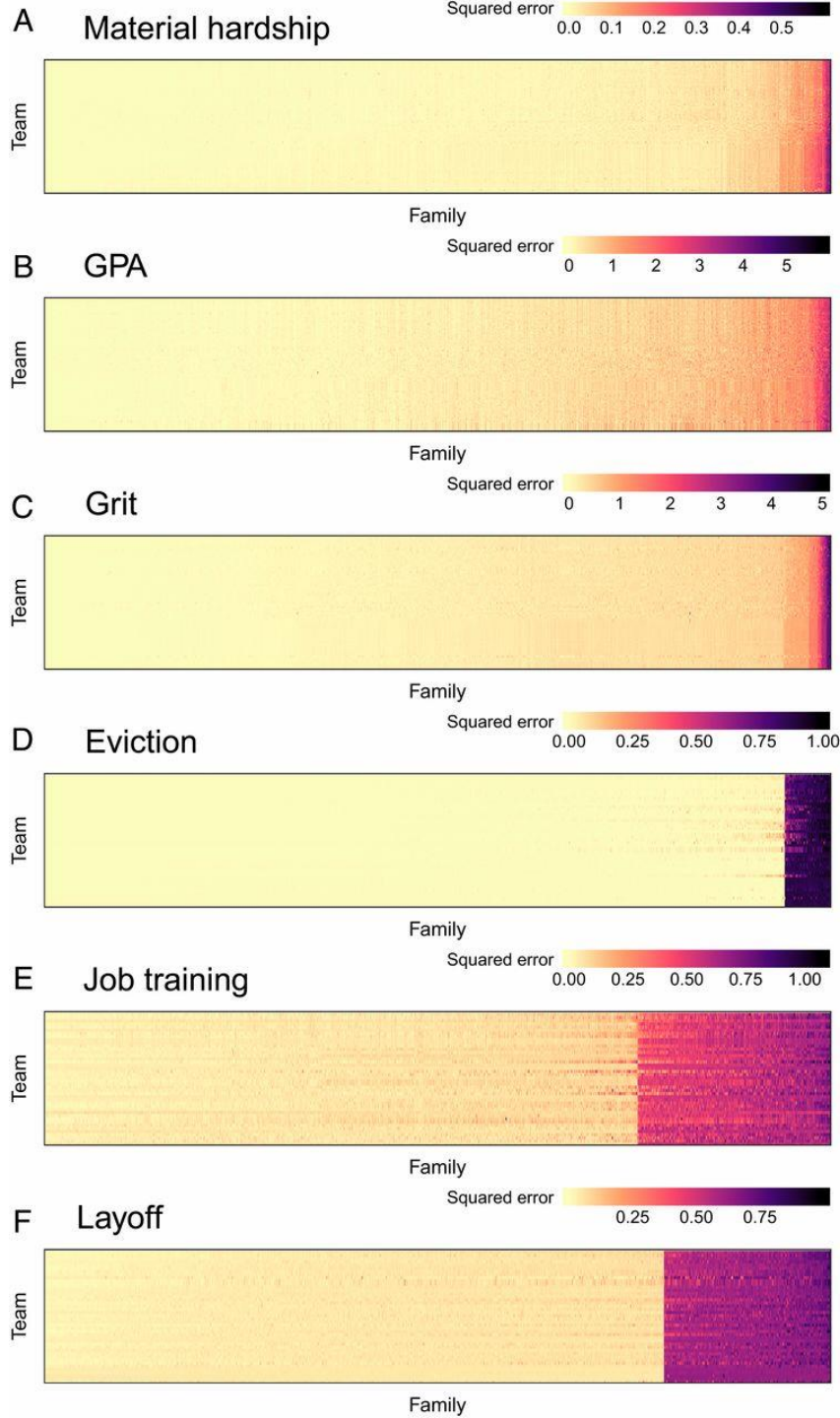


Image credit: Matthew Salganik, via Arvind Narayanan talk here: [MIT-STS-AI-snakeoil.pdf (princeton.edu)]

Additional empirical findings

- Teams used a huge variety of statistical/machine learning models, but they essentially all performed similarly: *"For all outcomes, the distance between the most divergent submissions was less than the distance between the best submission and the truth"*
- Some *families* were correctly predicted by all models, other *families* were incorrectly predicted by nearly everyone



The *Fragile Families* study shows us that predicting life outcomes is hard.

And the fanciest machine learning isn't better than the simplest models

Why is it hard?

- When will machine learning methods fail?
- Why is it different than, e.g., playing Go or Chess?

Progress in machine learning

Substantial progress

- Shazam, reverse image search
- Med. diagnosis from scans
- Speech to text
- Deepfakes
- Game playing machines: Chess, Go, Video games, etc.

Objective ground truth, "perception" or "reverse prediction"

Some progress

- Spam detection
- Copyright violation
- Automated essay grading
- Hate speech detection
- Content recommendation

Guessing (average) human judgement

Fundamentally dubious

- Predicting recidivism
- Predicting job success
- Predictive policing
- Predicting terrorist risk
- Predicting at-risk kids
- Predicting health outcomes

Predicting the future

Most prediction tasks in which machine learning does well, is NOT about predicting *the future!*

Original Content credit: Arvind Narayanan talk here: [MIT-STS-AI-snakeoil.pdf (princeton.edu)]

What makes prediction hard? (An incomplete list)

Methodological (data, computational, hardware, algorithmic) limitations

- Would be convenient if this was the answer – likely will be overcome eventually
- For many tasks (see last slide), this has been and is the answer

Measurement/data challenges – don't have the right data

- Slightly less convenient if this is the answer
- Maybe overcome, but likely need to make trade-offs with privacy, cost

Fundamental limits (universal noise/nondeterminism/irreducible noise)

- Very inconvenient answer (for researchers/companies) -- simply can't succeed
- Claim: many predictions with/about people is in this category

Original Source: Class on "Limits to prediction," by Arvind Narayanan and Matt Salganik, Princeton

Why do limits to prediction arise? (some hypotheses)

"Stochastics" – outcomes are truly random

- "Butterfly effect" – large outcome changes to small input changes
- Shocks – You can't predict who will win the lottery

Fundamental "data" issues

- Can never hope to collect some kinds of data
- "8 Billion problem" – number of training samples limited
- Drift – world is changing, so is mapping from data to outcome
- Disagreement on label, e.g., human judgement

Incentives, feedback loops, strategic behavior, and "equilibration"

Original Source: Class on "Limits to prediction," by Arvind Narayanan and Matt Salganik, Princeton

Other examples

- Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market
- Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices
- Predicting armed conflict: Time to adjust our expectations?
- Measuring the Effects of Advertising: The Digital Frontier.
- “Exploring Limits to Prediction in Complex Social Systems”
- “Can cascades be predicted?”
- Under the hood: Suicide prevention tools powered by AI.
- Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies.
- “The parable of Google Flu: Traps in big data analysis”,

Original Source: Class on “Limits to prediction,” by Arvind Narayanan and Matt Salganik, Princeton

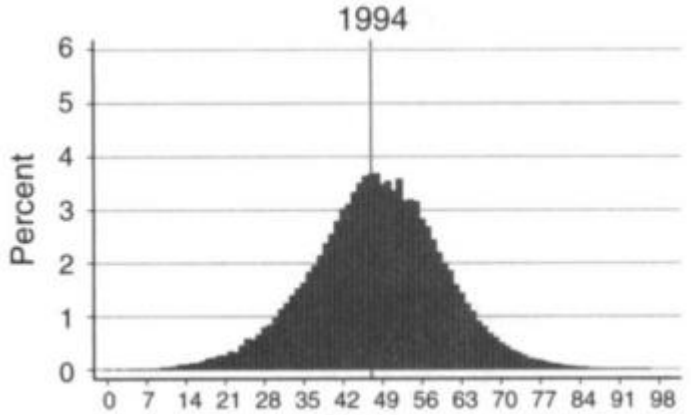
Incentives, feedback loops, strategic behavior, “equilibration”, and selection

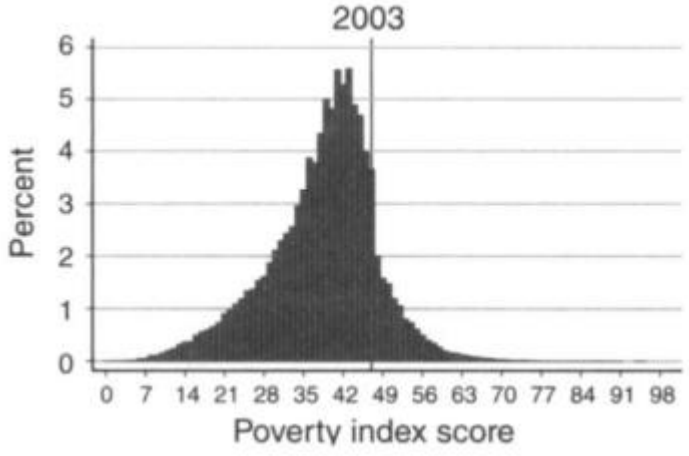
Difficulties and impossibilities when the prediction task is a player in the game

Example 1: Social program eligibility

In early 1990s Colombia, a social program was launched with eligibility below a certain “poverty score” threshold

Over time, people started gaming the poverty score to receive benefits





Original Source: “Manipulation of Social Program Eligibility,” Adriana Camacho and Emily Conover

I heard of this example via a talk by Tijana Zrnic, UC Berkeley

Example 2: Zillow

TECH

Zillow plunges 25% to lowest since July 2020, after company exits home-buying business

PUBLISHED WED, NOV 3 2021•1:33 PM EDT | UPDATED WED, NOV 3 2021•6:53 PM EDT

Ari Levy
@LEVYNEWS

SHARE

What happened?

Zillow's old business model, simplified

- Potential seller puts their home address onto Zillow
- Zillow predicts how much a house will be worth in the future
- Zillow subtracts a profit margin from that value
- Zillow makes an offer for the house
- Seller decides whether to accept or reject the offer

Claim:

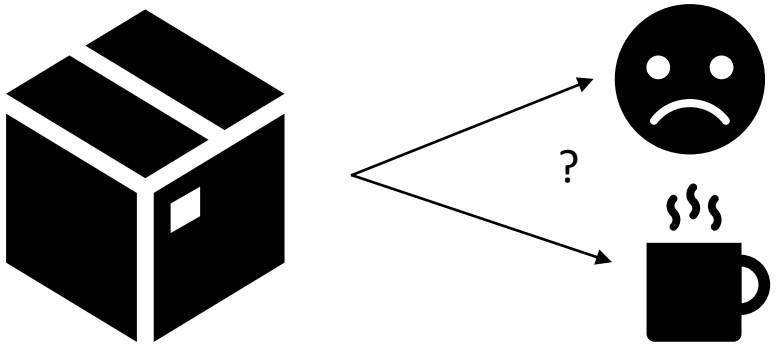
Main Idea – single seller

Suppose I have a box

With equal probability, either:

- Nothing
- Cup of coffee

I know what's in the box, but you don't



Coffee is worth

$1 to me

$1.50 to you

Does a sale happen?

If I can guarantee coffee, you pay me somewhere in [$1, $1.50] and we're both happy

If I can't, you offer me $0.75…Do I say yes?

Knowing the above, would you ever offer me $0.75?

NO! If you're smart… but a naïve machine learning would offer $0.75 anyway

Predictive models with strategic behavior

- In the Zillow example, local agents (potential sellers) are the ones who can decide whether to sell the box (house) or not, *after* they see the price offered by Zillow and the quality of the house
- Even if the model is correct on average, sellers being strategic mean that Zillow only gets to buy the houses in which they're wrong
- Possible to mitigate impact of such strategic behavior
  - Lily Xu's guest lecture: must modify machine learning model to consider poacher's reactions (changing where they poach)
  - Chess/Go algorithmic players: consider opponent's future play
  - Ongoing machine learning research: "Strategic classification"
- …But not eliminate impact
  - No model will ever catch 100% of poaching behavior

Conclusion

Questions to ask about a prediction task

Is this task about predicting the future?

- What is the fundamental noise (luck) in the outcome to be predicted?
- Do you believe that the features you have tell you a lot about the outcome?

Is this a task that a trained human expert would be good at?

- If not, why are humans bad at this task?
  - Why could a machine overcome these challenges?
- If so, what procedures does a human use, and can an algorithm hope to mimic or replace those procedures?

Will the future look like the training data?

- Distribution shift
- Strategic reactions to your model

AI is not magic: don't just appeal to "fancy algorithms" as solving prediction tasks – many domains in which they fail!

Announcements

- Part 1 of project due tomorrow night – NO LATE DAYS
- Regular office hours today (Zhi)
- No office hours Wednesday or Friday this week (Thanksgiving)
  - No class Wednesday
- Back to standard in-person lectures on Monday
- Details on turning in Project Part 2 posted soon
  - For now, make sure your code can run on Google Colaboratory with default packages
  - Cannot take more than ½ seconds per customer
  - Code + any needed data (e.g., a pickled trained model can be stored) in less than 10MB