MACHINE LEARNING & PUBLIC POLICY

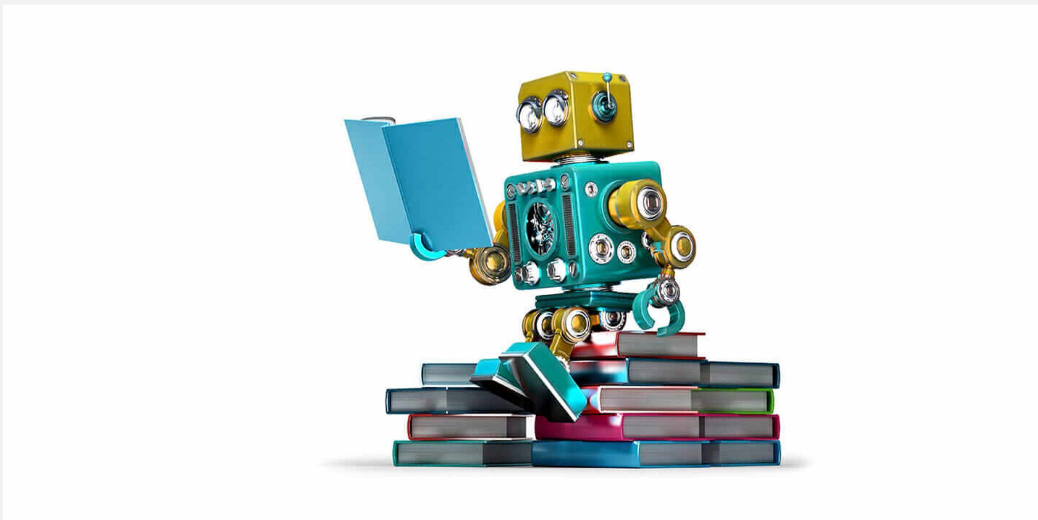
LECTURE 1: PREDICTION FOR POLICY

Professor Edward McFowland III

Information Systems and Decision Sciences
Carlson School of Managment

University of Minnesota

WHAT IS MACHINE LEARNING?



WHAT IS MACHINE LEARNING?

- **Machine Learning (ML)** is the study of systems that improve their performance at a task with experience (typically by **learning** from data)
- Some related concepts include
  - **Artificial Intelligence (AI)** is the science of automating complex behaviors such as learning, problem solving, and decision making.
  - **Data Mining (DM)** is the process o f extracting useful information from massive quantities of complex data.
  - **Statistics** is a branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data.

WHAT IS ML GOOD FOR?

WHAT IS ML GOOD FOR?

- Modeling Data and Learning Functions
  - Prediction/Classification
  - Grouping Data
  - Uncovering Correlations and Patterns

WHAT IS PUBLIC POLICY?

WHAT IS PUBLIC POLICY?

- **Public Policy (PP)** is the means by which a government maintains order or addresses the needs of its citizens
- A related concept is **Economics (Econ) which is the study of**
  - Scarcity, how people use resources and respond to incentives, and decision making
  - Agents making ("optimal") decisions, given an objective function, under constraints

ARE THESE TWO EVEN RELATED?

- **Public Policy** has traditionally placed emphasis on the understanding and explanation of (causal) relationships
  - With the usual (active) goal to make beneficial policy decisions
- **Machine Learning** has traditionally placed emphasis on discovering patterns of (correlational) relationships
  - With the usual (passive) goal of providing an estimate of some target outcome

ARE THESE TWO EVEN RELATED?

- ML and Policy questions both begin with some notion of $Y = f(X) + \epsilon$
  - Y is an outcome variable (GPA), X is an input variable (Tutoring Program), ϵ is just random noise
  - There exist a relationship (measure by f) between X and Y, that we want to estimate "well" (f)
- True data generating process (DGP) can be a subset of the following:
  - Getting tutoring (causally) improves the GPA attained $(X \rightarrow Y)$
  - GPA (causally) changes how much tutoring is received $(X \leftarrow Y)$
  - Inherit work ethic (Z), which is contained in ϵ as an omitted variable, (causally) changes both amount of tutoring received and the GPA attained $(X \leftarrow Z \rightarrow Y)$
- Different ways to measure estimation quality, e.g.,
  - $R(f,f) = E\left[L(f,f)\right] \stackrel{\text{def}}{=} \int_x \left(f(x) - f(x)\right)^2 p(x)dx$
- Can estimate f from data $(y_i, x_i)_{i=1\ldots n}$ well as there is correlation between X and Y
- Generally we cannot identify what is the true DGP (i.e., causal relationship)
  - Sufficient for using X to predict Y
  - Insufficient for using X as a decision variable to causally to change Y

HAVE WE REACHED AN IMPASSE?

- **Observation I:** Sometimes correlation is valuable on its own
- **Observation II:** Sometimes goal is one of detection or discovery
- **Observation III:** Sometimes correlation can be (forced into) causation

PREDICTION FOR POLICY

Observation 1: Sometimes correlation is valuable on its own

MOTIVATION: AN EDUCATION EXAMPLE

LARGE CLASSROOMS VS SMALL CLASSROOMS

MOTIVATION: AN EDUCATION EXAMPLE

MOTIVATION: AN EDUCATION EXAMPLE

CLASSROOM SIZE VS TEACHER TURNOVER

PUBLIC POLICY: A STORY PREDICTION AND CAUSALITY

- There are many critical question that are challenges of prediction
  - Ex: Which teacher(s) will churn?
- Predictions are (often implicitly) key variables into critical policy decisions
  - Ex: How many teachers to recruit next year—which impacts budgets, transfers, marketing, etc—depends on predicting teacher churn.
- Prediction can lead to causation
  - Ex: Once I know which teachers are leaving, I may also want to intervene and incentive good teachers to stay.
- Policy makers must "learn" to "classify" problems as those of prediction

PREDICTION FOR POLICY: TRAINING DATA

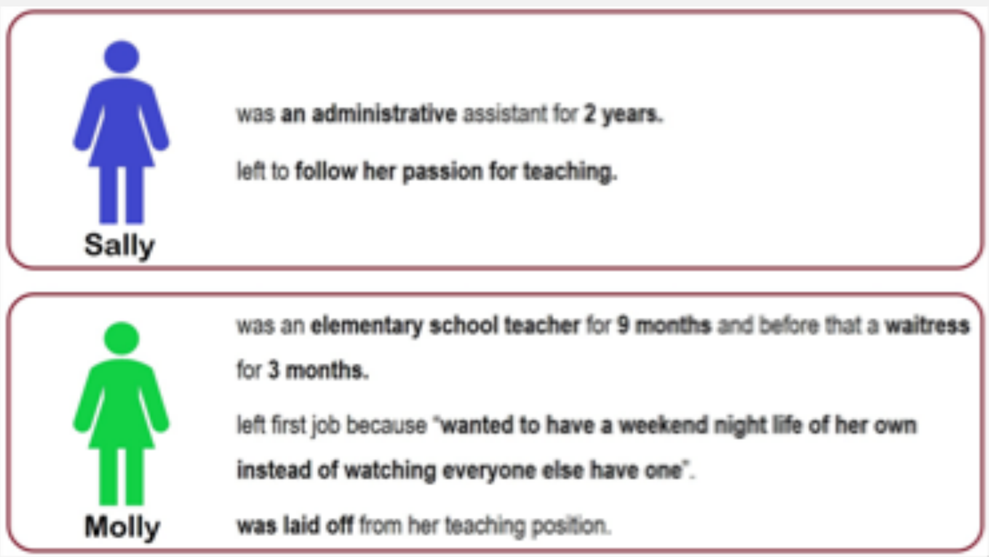
Observation 1: Sometimes correlation is valuable on its own

THE WORLD'S BIGGEST CHALLENGES

The 2030 Agenda for Sustainable Development, United Nations, 2015

HIRING IN EDUCATION

WHO DO YOU HIRE?



CAN ML HELP?

- Discuss in Groups.

1. Characterize how you think this is typically accomplished?
2. Can this be structured as a challenge of prediction? How?
3. What kind of data would be necessary? Useful?
4. What are the features and outcomes of interests?
5. What benefits does ML offer?

S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

HIRING IN EDUCATION

- **Research Question:** to what extent can resumes be used to generate a cheap, accurate, fair, credible signal to improve teacher hiring decisions?
- **Approach:**
  - Link applicants' resumes to effectiveness & retention as hires
  - Create theory-informed predictor variables from resume data
    - It is critical for policy challenges to include policy experts
  - Evaluate prediction-model's value

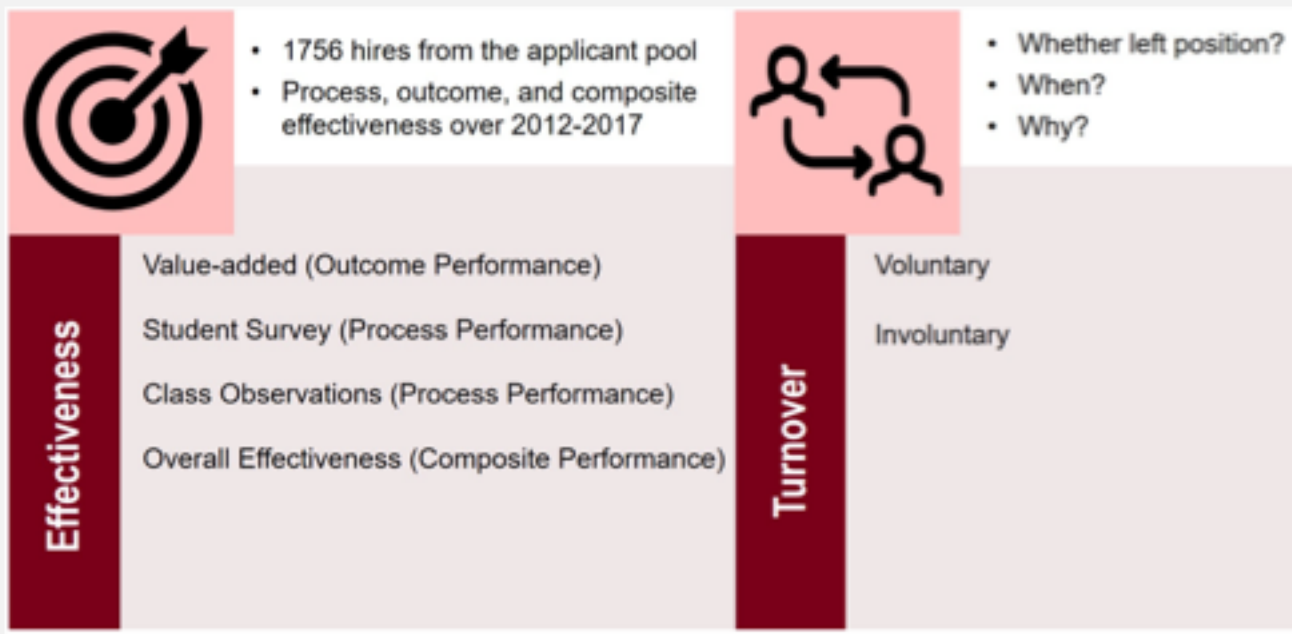
S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

HIRING IN EDUCATION

APPLICANT INFORMATION					
Last Name		First		Date	
Gender		Ethnicity			
Position Applied for					
Have you ever worked for this district?	YES ☐ NO ☐				
EDUCATION					
School		Degree			
From		To		Major	
PREVIOUS EMPLOYMENT					
Company		Job Title			
Supervisor		Phone			
Job Description					
From		To			
Reason for Leaving					

S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

HIRING IN EDUCATION



S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

CRITICAL FEATURES

- Work Experience Relevance
  - Positive association with teacher performance
  - Negative association with turnover hazard
- Previous Tenure
  - Positive association with teacher performance
  - Negative association with turnover hazard
- Turnover History
  - Involuntary
    - Positive association with teacher performance and negative with turnover hazard
  - Avoiding bad jobs
    - Negative association with teacher performance and negative with turnover hazard
  - Approaching better jobs
    - Positive association with teacher performance and negative with turnover hazard

S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

CRITICAL FEATURES

- Work Experience Relevance
  - Job description → Occupation Code → Occupation Dimensions → Relevant Experience

S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

CRITICAL FEATURES

- Work Experience Relevance
  - Job description → Occupation Code → Occupation Dimensions → Relevant Experience
  - Naïve Bayes trained on occupation descriptions, used to classify job titles and descriptions into Occupation Codes.

S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

CRITICAL FEATURES

O*NET						
O*NET Code	O*NET Occupation	Knowledge	Skills	Abilities	Styles	Interests
35-3031.00	Waiters and Waitresses					
43-6014.00	Secretaries & Administrative Assistants					
41-4012.00	Sales Representatives					

Applied to Elementary School Teacher Position						
O*NET Code	O*NET Occupation title	Knowledge	Skills	Abilities	Styles	Interests
25-2021.00	Elementary School Teachers					

- Work Experience Relevance
  - Job description → Occupation Code → Occupation Dimensions → Relevant Experience
  - Naïve Bayes trained on occupation descriptions, used to classify job titles and descriptions into Occupation Codes.

S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

CRITICAL FEATURES

Involuntary Turnover	Job Dissatisfaction	Following Passion
Layoff	Dissatis(fied)(faction)	Love
Budget cut	Exhaust(ed)(ing)	Passion
Excessed	Condition	Dream
Reduction	Commute	Excite
Let go	Environment	Teach

- Work Experience Relevance
  - Job description → Occupation Code → Occupation Dimensions → Relevant Experience
  - Naïve Bayes trained on occupation descriptions, used to classify job titles and descriptions into Occupation Codes.
- Turnover History
  - Take a small sample, and hand code reasons for turnover (Involuntary, Avoiding bad jobs, Approaching better jobs)
  - Naïve Bayes trained on these labeled examples to predict turnover reason.

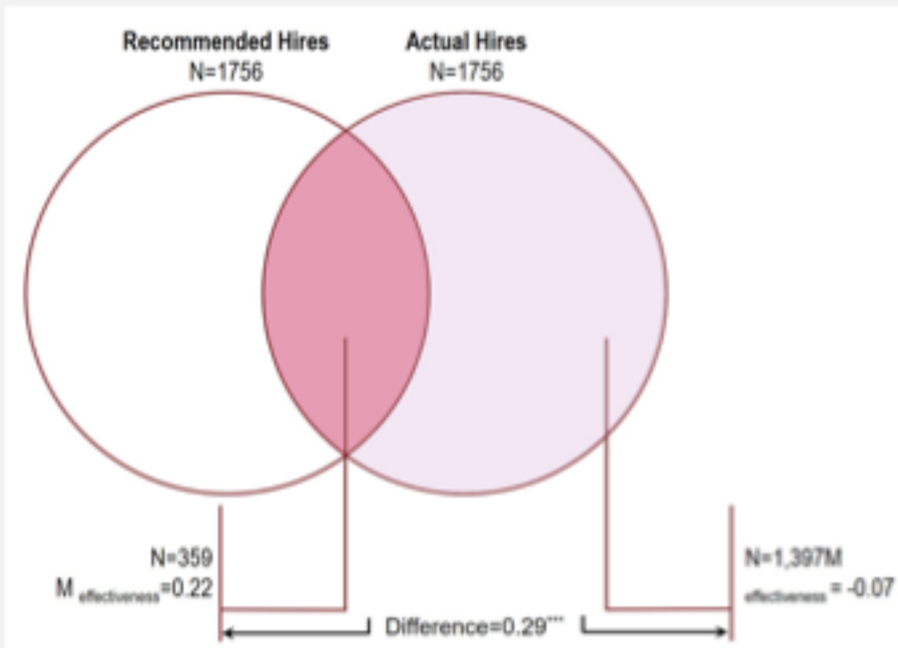
S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

OTHER FEATURES

- Tone (Linguistic Inquiry and Word Count Software, Pennebaker et al., 2015)
- Spelling accuracy
- Propensity for job-hopping
- Teaching experience
- Education
- Number of previous jobs
- Overall work experience
- Unemployment gaps
- Application year
- Position type
- Demographics (missing data imputed)
- Quality and Quantity of competition faced by applicant (used for IV)

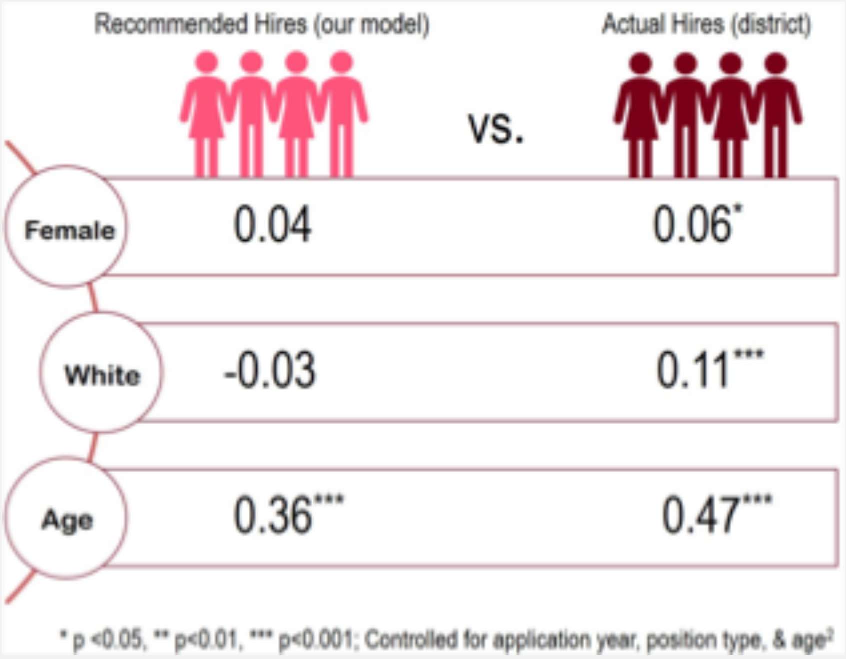
S. Sajjadiani et al. "Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover," (Working Paper).

RESULTS



Recommended Hires
N=1756
Actual Hires
N=1756
N=359
M effectiveness=0.22
N=1,397M
effectiveness = -0.07
Difference=0.29***

RESULTS

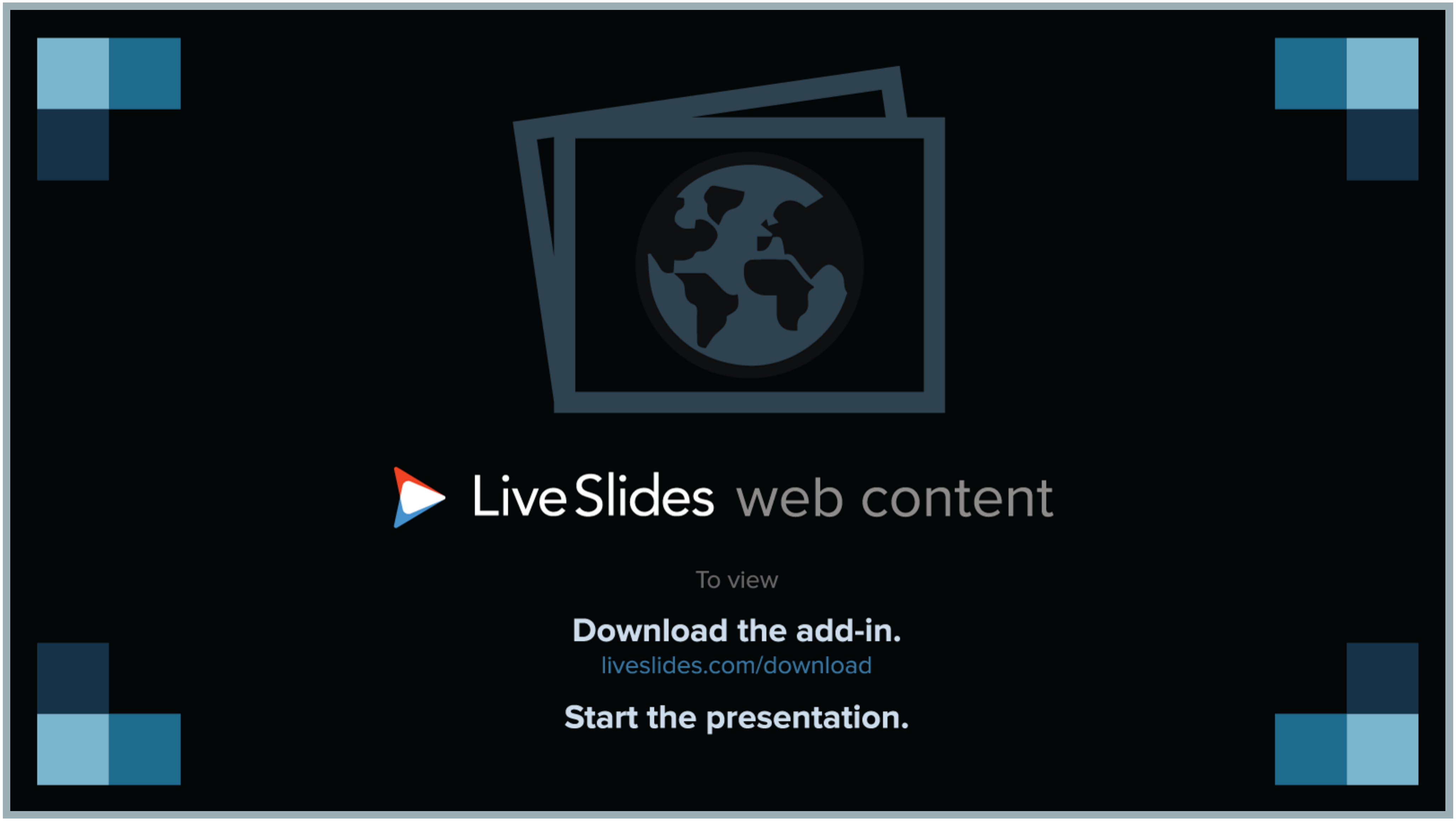


Recommended Hires (our model)
Actual Hires (district)
vs.
Female
0.04
0.06*
White
-0.03
0.11***
Age
0.36***
0.47***
* p <0.05, ** p<0.01, *** p<0.001; Controlled for application year, position type, & age2

THE WORLD'S BIGGEST CHALLENGES

POVERTY

- Eliminating Poverty is critical UN Sustainable Development Goal
- To address this problem we first need to have accurate measure of poverty by location
  - Traditional surveys are expensive and inefficient
- How can ML help capture poverty in various areas?
- Discuss in groups.
  1. Can this be structured as a challenge of prediction? How?
  2. What kind of data would be necessary? Useful?
  3. What are the features and outcomes of interests?
  4. What benefits does ML offer?

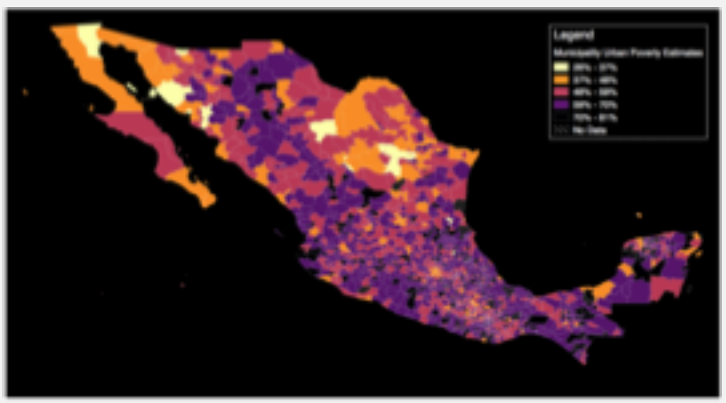

LiveSlides web content
To view
Download the add-in.
liveslides.com/download
Start the presentation.

POVERTY

- What were the challenges?
  - Only a small number poverty level labels are known
  - Unclear which features are important to predict poverty levels of an area
  - Inefficient/Costly to conduct surveys to solve these issues
- How were they overcome?
  - Poverty levels highly correlated with different features of economic development
    - Economic development can be captured through high resolution images
  - Poverty levels (and therefore economic development) also highly correlated with intensity of light levels at night
  - Learn how night light intensity predicts poverty levels in the locations where labels are known
  - Use algorithms to learn "features" in daytime images that predict night light intensity
    - Infrastructure, Roads, Amount of Urban Areas/Farmland, Waterways etc.
- Final Results: Day images feature → Night light intensity → Poverty level

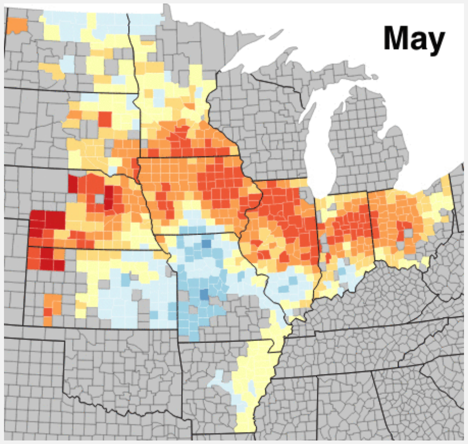
N. Jean et al. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794.

POVERTY



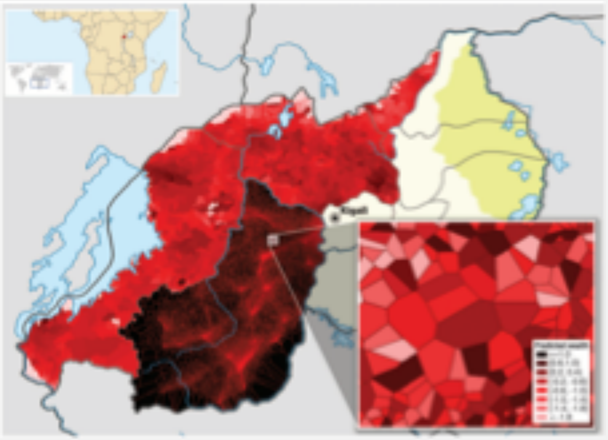
- Does this work in other countries?
  - Babenko et al. does something similar in Mexico

POVERTY



- Does this work in other countries?
  - Babenko et al. does something similar in Mexico
- Does this work for other contexts?
  - You et al. does something similar for predicting crop yields.

POVERTY



- Does this work in other countries?
  - Babenko et al. does something similar in Mexico
- Does this work for other contexts?
  - You et al. does something similar for predicting crop yields.
- Can we get more granular?
  - Blumenstock et al. predicts individual poverty levels based on historical cell phone call records.

THE WORLD'S BIGGEST CHALLENGES

INCLUSIVE FINANCE

2-3 BILLION individuals and **200 MILLION** businesses in emerging economies today lack access to savings and credit, and even those with access can pay dearly for a limited range of products.

McKinsey Global Institute, Digital finance for all: Powering inclusive growth in emerging economies, 2016

INCLUSIVE FINANCE

Goal 8.10 of the "SDGs"

Strengthen the capacity of domestic financial institutions to encourage and expand access to banking, insurance and financial services for all.

McKinsey Global Institute, Digital finance for all: Powering inclusive growth in emerging economies, 2016

INCLUSIVE FINANCE

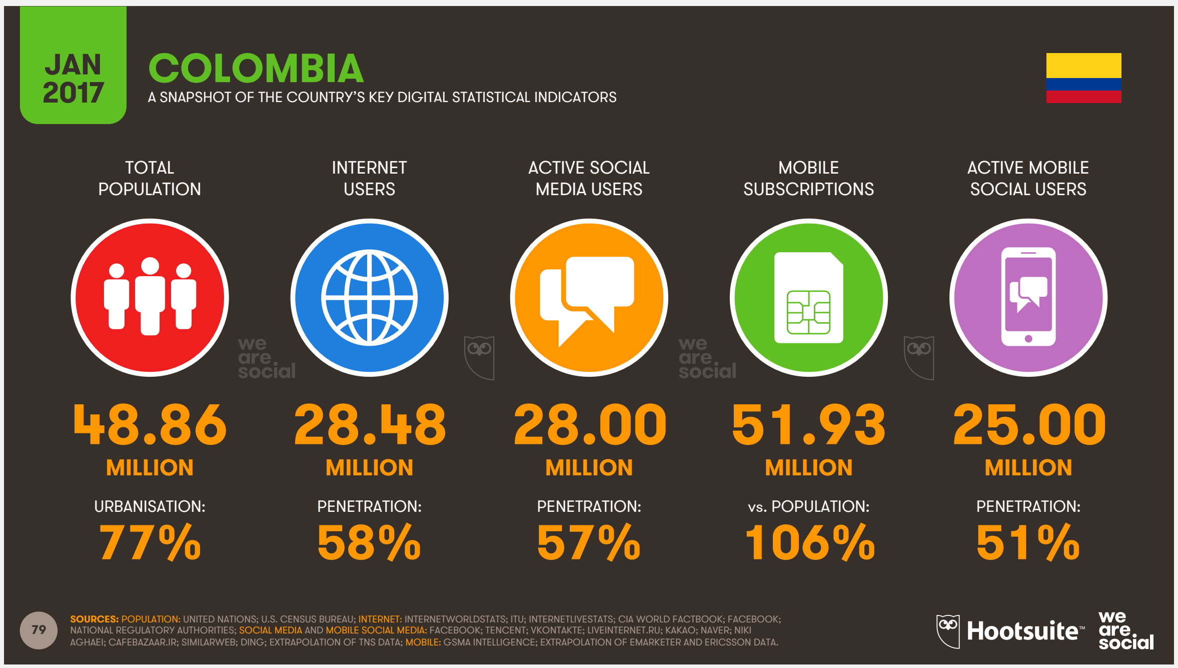
Banking Penetration in Colombia

2014	# Individuals	% Adult population
Savings account	21.6 million	68%
Credit product	7.3 million	23%
Credit card	7.0 million	22%
Checking account	1.5 million	5%
Mortgage	0.9 million	3%
Total population	48.0 million	32.0 million

Source: Bancolombia

- Access to credit can be extremely valuable
- Lenders face two challenges with unbanked population
  - How to evaluate the credit worthiness of a loan seeker
  - How to transact with a loan seeker
- Discuss in groups.
  1. Can this be structured as a challenge of prediction? How?
  2. What kind of data would be necessary? Useful?
  3. What are the features and outcomes of interests?
  4. What benefits does ML offer?

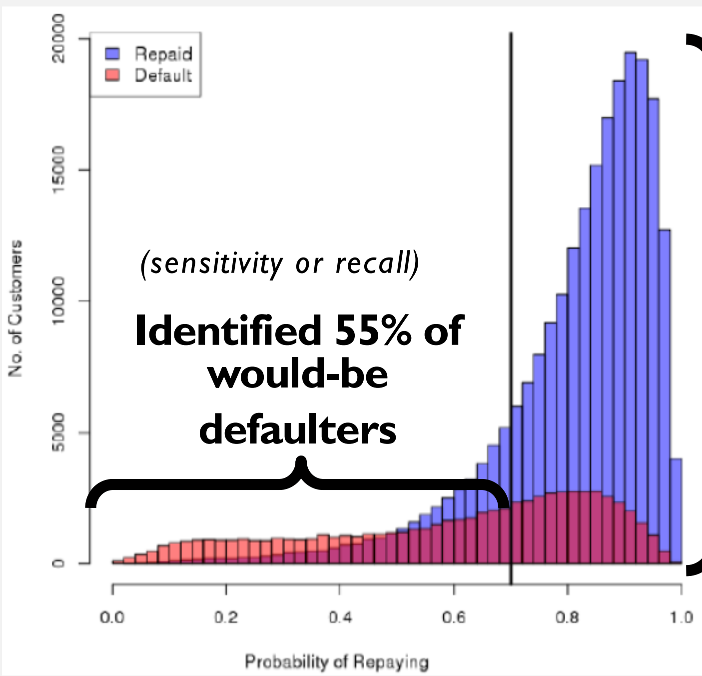
INCLUSIVE FINANCE



INCLUSIVE FINANCE

- Mobile Money platforms provide a coinvent way to send/receive money over mobile phones
  - Under-banked populations now have access to previously unavailable credit
- Mobile network usage data provides a rich set of features
  - From Airtime usage patterns to the credit-worthiness of network “friends”

INCLUSIVE FINANCE

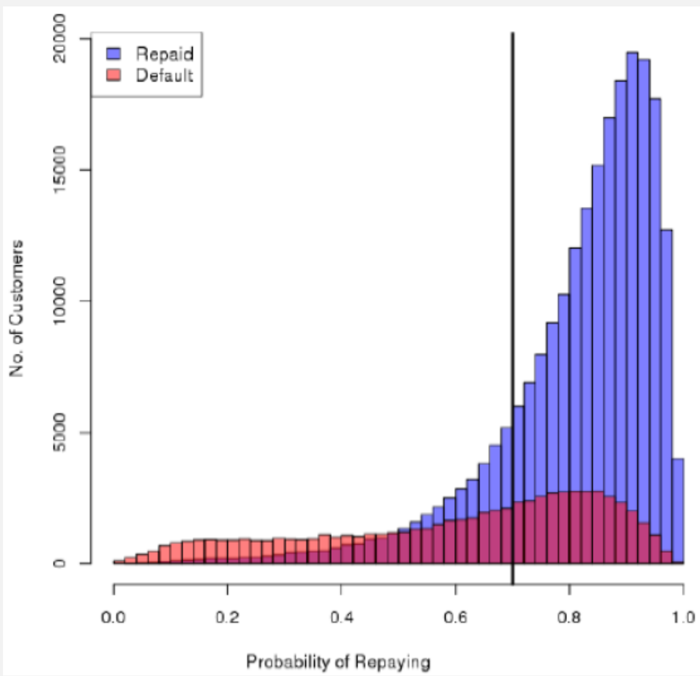


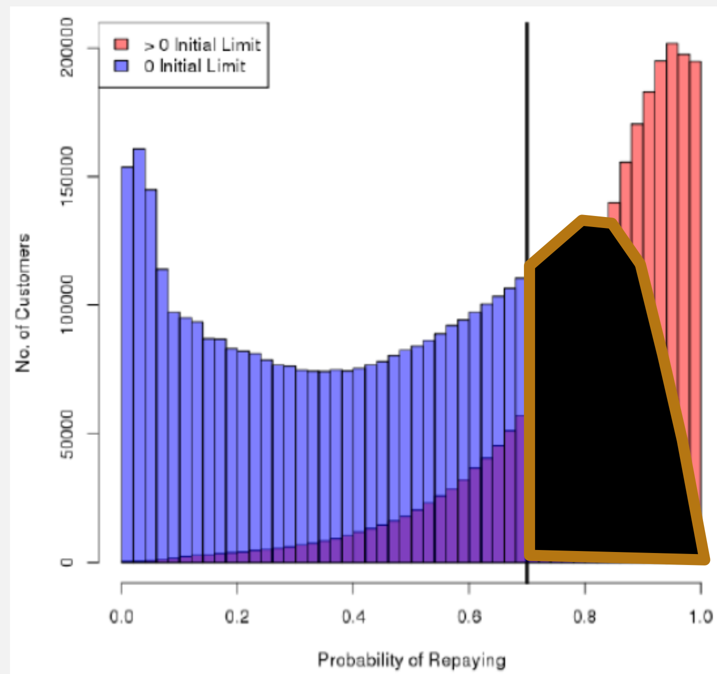
(specificity)

Retained 83% of paying customers

Speakman

INCLUSIVE FINANCE





+1 Million customers receive credit

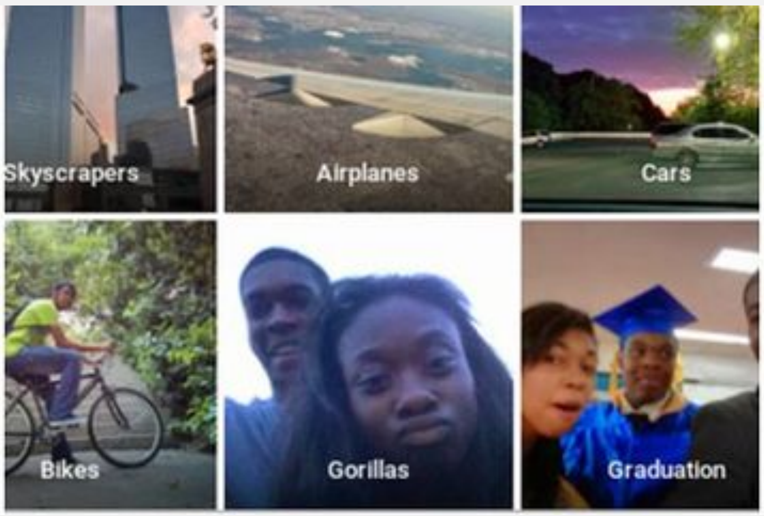
Speakman

THE WORLD'S BIGGEST CHALLENGES

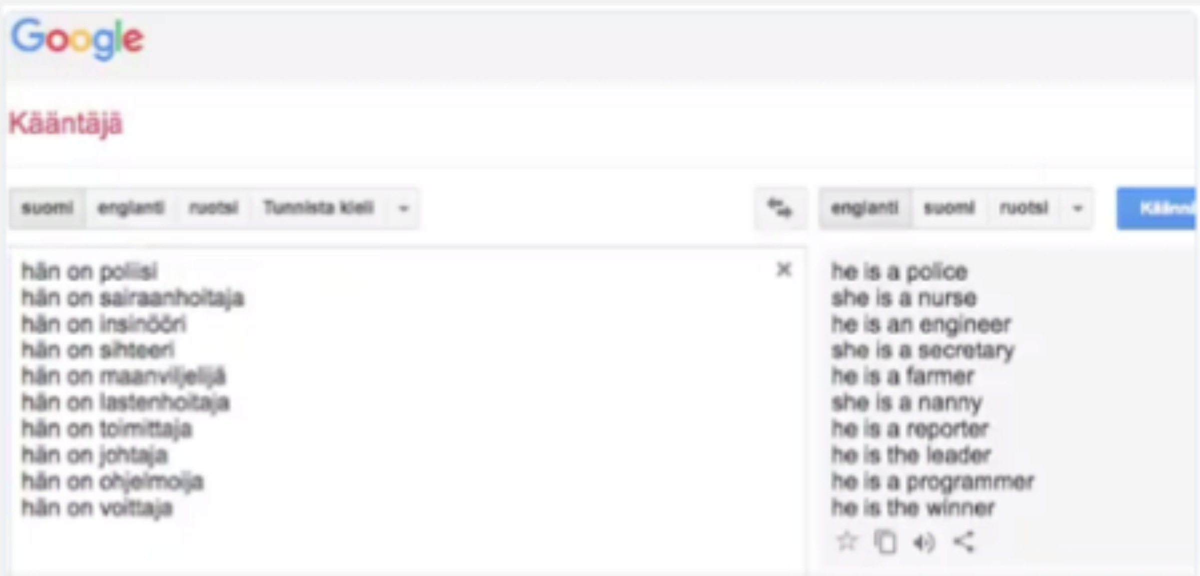
THE IMPORTANCE OF FATE

- Fairness, Accountability, Transparency, and Ethics (FATE)
- We continue to rely more and more on algorithms to decide who gets
  - hired, a loan, parole from jail, etc.
- We have to insure that the algorithm's decisions are not biased
- People trust models because they "appear" to remove human bias but this does not mean they are devoid of algorithmic bias
  - The machine is learning, but we provide the textbook to teach it.

THE IMPORTANCE OF FATE



THE IMPORTANCE OF FATE



LiveSlides web content

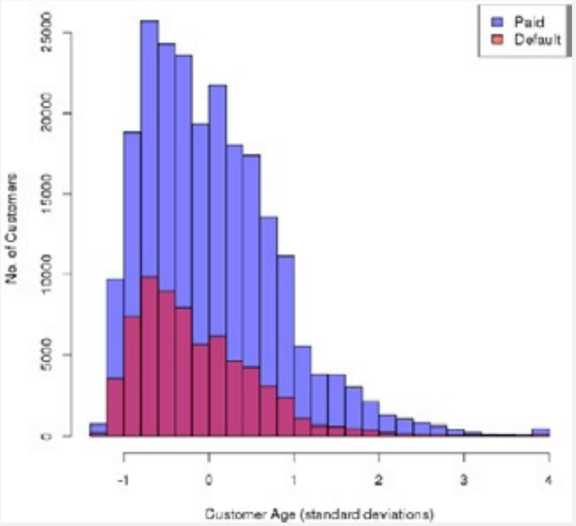
To view

Download the add-in.

liveslides.com/download

Start the presentation.

THE IMPORTANCE OF FATE



THE IMPORTANCE OF FATE

PREDICTION FOR POLICY

Observation 1: Sometimes correlation is valuable on its own

REFERENCES

- Sima Sajjadiani, Aaron J. Sojourner, John D. Kammeyer-Mueller & Elton Mykerezi. (2018). Using Machine Learning to Translate Applicant Work History into Predictors of Performance and Turnover. Working Paper.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794. http://doi.org/10.1126/science.aaf7894
- Boris Babenko, Jonathan Hersh, David Newhouse, Anusha Ramakrishnan, Tom Swartz. (2017). Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, With an Application in Mexico. arXiv.org.
- Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, Stefano Ermon, Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data, 31th AAAI Conference on Artificial Intelligence.
- Blumenstock, JE, Cadamuro, G, On, R (2015). Predicting Poverty and Wealth from Mobile Phone Metadata, Science, 350(6264), 1073-1076.
- ProPublica. Machine Bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

ADDITIONAL MATERIALS

- ML and Policy
  - Machine Intelligence and Public Policy (Sendhil Mullainathan)
    - https://youtu.be/W7izsrYGhdU
  - Economics in the age of big data (Liran Einav and Jonathan Levin)
    - http://web.stanford.edu/~leinav/pubs/Science2014.pdf
  - Big Data: New Tricks for Econometrics (Hal Varian)
    - https://www.aeaweb.org/articles?id=10.1257/jep.28.2.3
    - https://web.stanford.edu/class/ee380/Abstracts/140129-slides-Machine-Learning-and-Econometrics.pdf
- Bias in ML
  - The Trouble with Bias (Kate Crawford)
    - https://youtu.be/fMym_BKWQzk