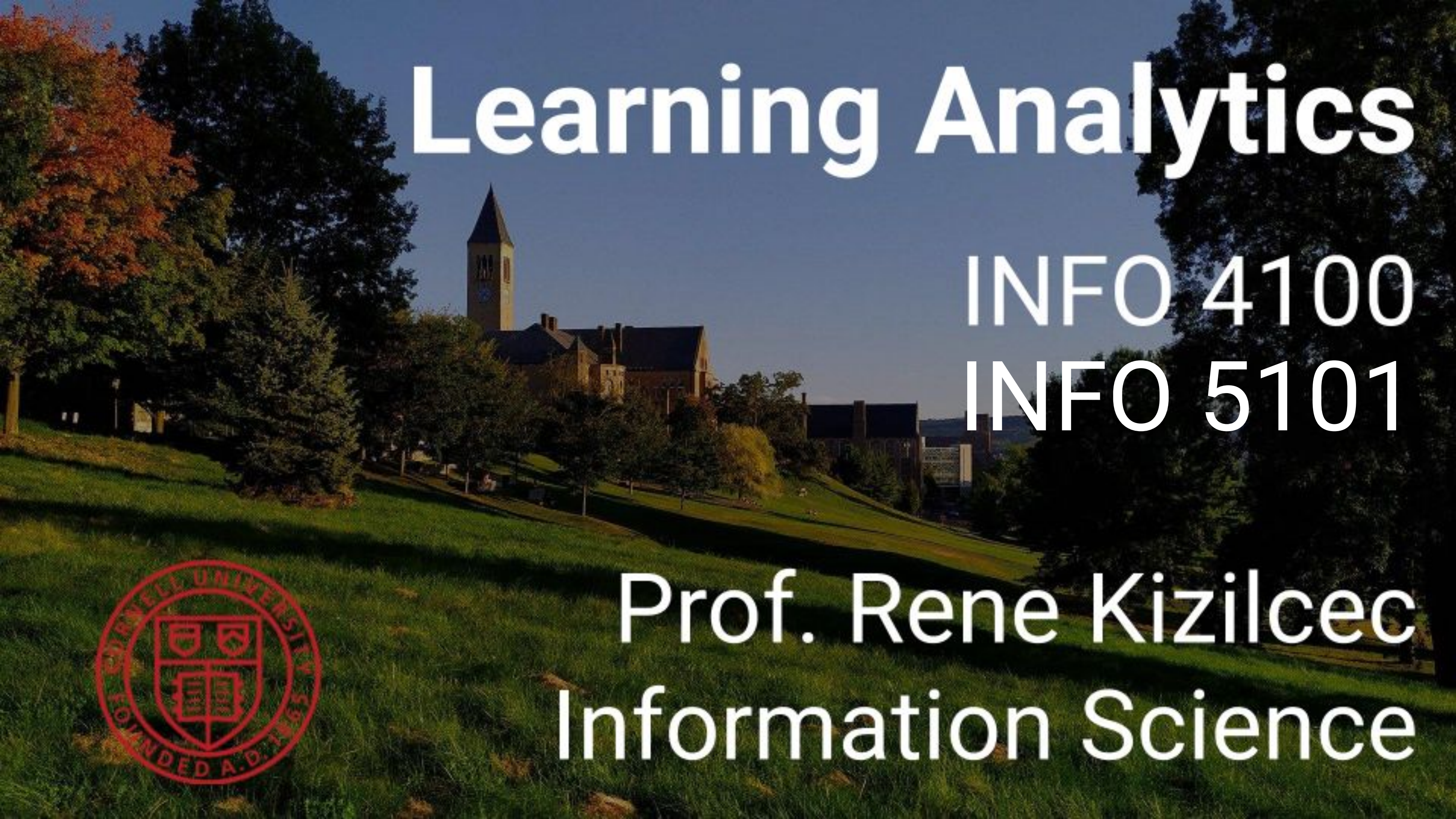


Learning Analytics

INFO 4100
INFO 5101

Prof. Rene Kizilcec
Information Science

A few logistics

- Need to have formed teams for first group project
- If you enrolled late, assignments due before joining will be discounted, other work will count more
- For extension requests, Slack DM your section TA and me
- HW Extensions: submit a placeholder before the deadline
- Knitting to Word requires Word desktop app, insert screenshots if needed
- Progress page starts at 0% and goes up - scary, I know
- HiTA quick demonstration (NB: smaller file, only *tidyverse*)

The Ethics of Learning Analytics

What is the object of ethical consideration?

The Ethics of Learning Analytics

What is the object of ethical consideration?

- Personal data collected, monitoring of individuals
- (Imperfect) Predictions of outcomes that inform decisions
- Variation in available resources as part of experiments

Student data

What student data do colleges and universities routinely collect and store?

- Enrollments
- Gpa
- Demographic data
- High school stats
- Employment outcomes

How could the data be used?

How could the data be used?

Some examples:

- Give feedback to learners / instructors
- Inform educational policy
  e.g. by revealing inequities in representation/achievement
- Evaluate effective learning / teaching practice
- Predict student outcomes for early intervention / adaptation

Emerging Themes from Reading Reflections

- Student Agency in Learning Analytics
- Privacy Concerns about Sensitive Data
- Potential Misuse and Misinterpretation of Data
- Critical Perspectives on Data and Neutrality
- Personalized Learning for Educational Enhancement

Four Principles of Moral Ethics

Autonomy

Beneficence

Non-maleficence

Distributive justice

Four Principles of Moral Ethics

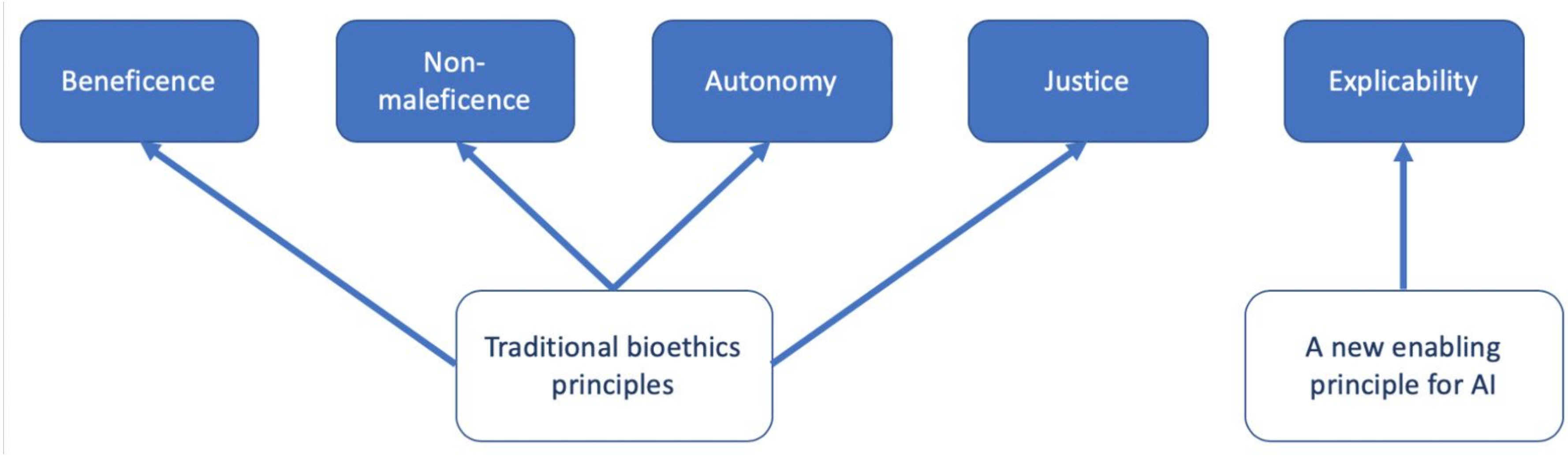
Autonomy		Availability and benefits to all
Beneficence	Which is which?	Intend to do good
Non-maleficence		Informed consent
Distributive justice		Do no harm

Basic Moral Principles

Autonomy	Availability and benefits to all
Beneficence	Intend to do good
Non-maleficence	Informed consent
Distributive justice	Do no harm

Explicability in AI decision-support systems

Comprehensive review concluded that ethical principles for AI in society also include **explicability** Foridi & Cowls, 2019



Explicability incorporates **intelligibility** (*how does it work?*)
and **accountability** (*who is responsible for how it works?*)

Learning analytics are not fair by default

- An unequal relationship: Student <--> Institution
- Providers act in “paternalistic”, “authoritarian” ways
- Surveillance and hidden algorithms
- Student consent and agency

Family Educational Rights and Privacy Act (FERPA)

Family Educational Rights and Privacy Act (FERPA)

FERPA gives students four basic rights with respect to their education record:

1. The right to inspect and review their education record.
2. The right to **<u>some</u>** control over the disclosure of their education record.
3. The right to request the amendment of their education record where appropriate.
4. The right to file a complaint with the U.S. Department of Education concerning alleged failures by the University to comply with the requirements of FERPA.

See https://registrar.cornell.edu/service-resources/ferpa

Two Discussion Activities

1. Focusing on Needs of Different Stakeholders

2. Focusing on Different Values and their Implications on Design

Different Stakeholders, Different Needs/Concerns

Splitting up the room into 3 groups:

Universities/Schools | Students | Teachers

You are tasked with writing a policy memo with 3-4 key points to represent your group's needs and concerns.

Spend 7 minutes talking to two students around you to develop your key points

Different Values, Different Designs

Splitting up your group room into 3 positions – pick one each:

Performance - **Equity** - **Privacy**

You are tasked with proposing how a system should be designed to achieve your value. Then the other 2 people in your group respond based on their value, raising concerns, or offering suggestions for consensus building

Spend 10 minutes talking to each other. Then share out your design desiderata

Let’s unpack

Thesis: We should collect a wide range of student data to predict which students are at risk of failing and allocate resources to help them succeed.

Predictive models: uses and biases

What should we (not) predict?

What variables should we (not) use in predictive models?

Questions for predictive models

- Which students will submit the homework late?
- What exercise to give to a student next?
- What affective state is the student experiencing?
- Which applicants will matriculate?
- Which students will complete their degree?
- Which students will donate later on?

Data collection and governance

What data should (not) be collected for learning analytics?

Should students have a ‘right to be forgotten’? How much? What if that means that future generations of students learn less b/c algorithms do worse?

What can we learn from college students' network transactions? Constructing useful features for student success prediction.

ABSTRACT

Identifying at-risk students at an early stage is a challenging task for colleges and universities. In this paper, we use students' on-campus network traffic volume to construct several useful features in predicting their first semester GPA. In particular, we build proxies for their attendance, class engagement, and out-of-class study hours based on their network traffic volume. We then test how much these network-based features can increase the performance of a model with only conventional features (e.g., demographics, high school GPA, standardized test scores, etc.). We labeled students as "above median" and "below median" students based on their first term GPA. Several machine learning models were then applied, ranging from logistic regression, SVM, and random forests, to AdaBoost. The result shows that the model with network-based features consistently outperforms the ones without, in terms of accuracy, f1 score, and AUC. Given that network activity data is readily available data in most colleges and universities, this study provides practical insights on how to build more powerful models to predict student success.

EDM 2018

Is it ethical for a university to use students' Internet network traffic (i.e. where they use a device and for how long)?

What if it improves at-risk student identification?

What if browsing history is even more predictive?

Where do you draw the line?

Keep these considerations in mind throughout the course.

BJET | British Journal of Educational Technology   BERA
BRITISH EDUCATIONAL RESEARCH ASSOCIATION

Original Article

Population validity for educational data mining models: A case study in affect detection

Jaclyn Ocumpaugh, Ryan Baker, Sujith Gowda, Neil Heffernan, Cristina Heffernan

The data used to develop EDM models are collected from students who may not be representative of the broader population. In order to use EDM models with new populations, their generalizability must be verified. **Do detectors of affect remain valid when applied to new populations?**

Models of four educationally relevant affective states were constructed based on data from urban, suburban and rural students using ASSISTments software for middle school mathematics in the Northeastern United States. **Affect detectors trained on a population drawn primarily from one demographic do not generalize to other demographics, even though they are part of the same national/regional culture.**

Models constructed using data from all three subpopulations are more applicable to students in those populations than those trained on a single group, but still do not generalize across all subgroups. Models generalize better across urban and suburban students than rural students.

Quantifying Bias

Recent method:

Absolute Between-ROC Area (ABROCA)

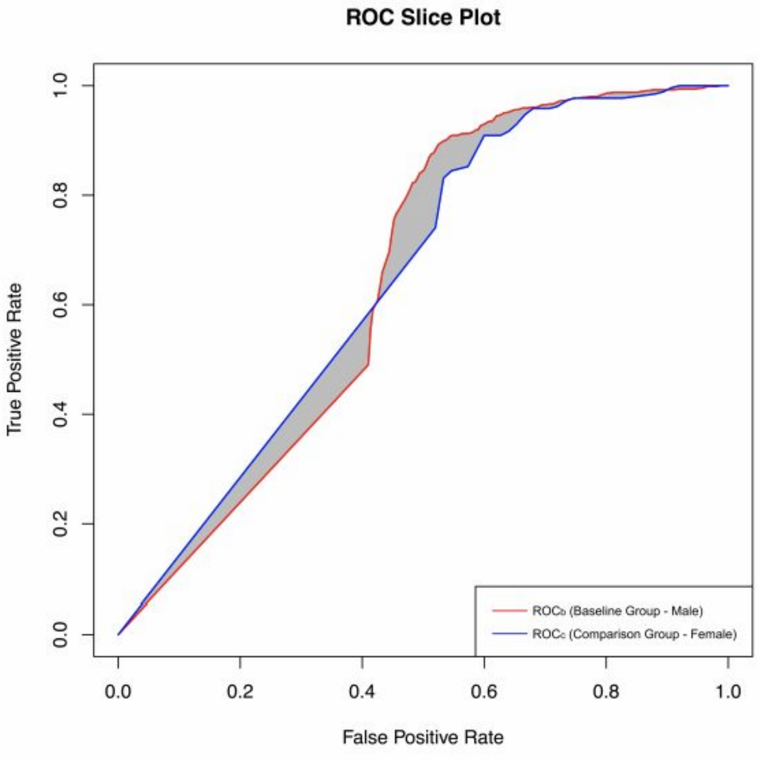
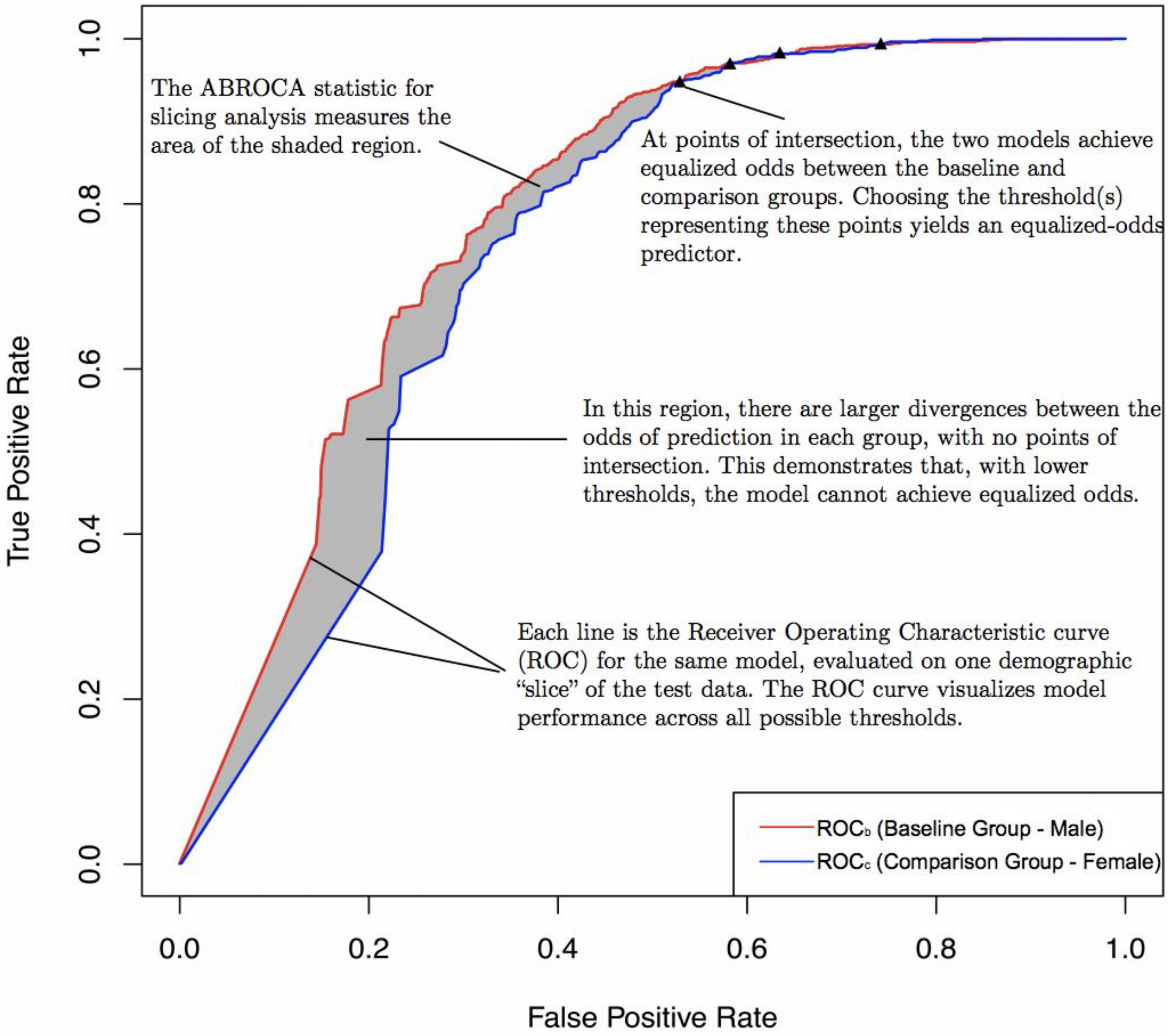


Figure 1: A "slice plot," which shows the Receiver Operating Characteristic (ROC) curves for a model across two groups (male and female learners). The shaded region between the two curves represents the ABROCA statistic, which measures differences in model performance across groups.

Gardner, Brooks, & Baker, LAK'19


The ABROCA statistic for slicing analysis measures the area of the shaded region.
At points of intersection, the two models achieve equalized odds between the baseline and comparison groups. Choosing the threshold(s) representing these points yields an equalized-odds predictor.
In this region, there are larger divergences between the odds of prediction in each group, with no points of intersection. This demonstrates that, with lower thresholds, the model cannot achieve equalized odds.
Each line is the Receiver Operating Characteristic curve (ROC) for the same model, evaluated on one demographic "slice" of the test data. The ROC curve visualizes model performance across all possible thresholds.
ROC_b (Baseline Group - Male)
ROC_c (Comparison Group - Female)
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0

Gender Bias in Predicting MOOC Dropout

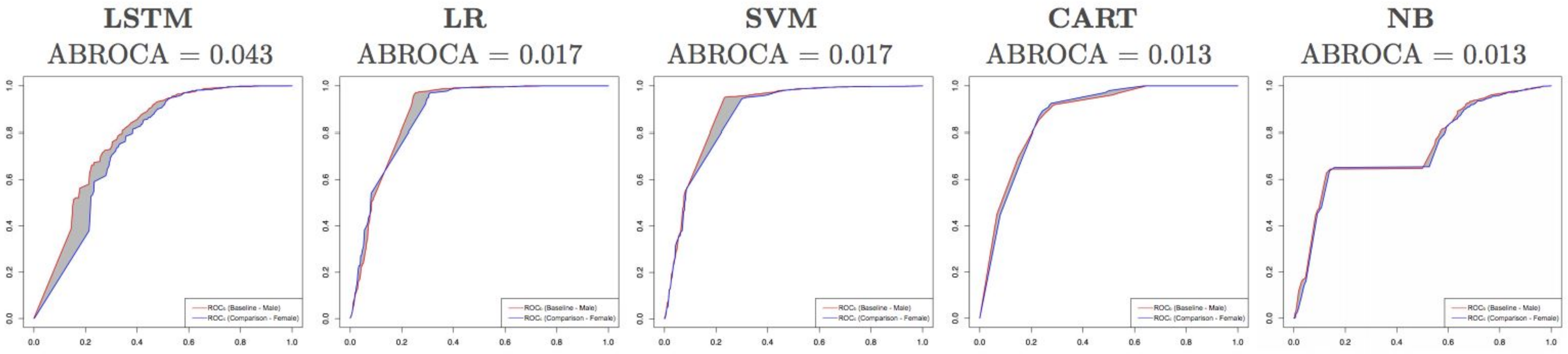


Figure 3: Slice plots for all models applied to the business MOOC used in Figure 2.

Let’s unpack

Thesis: We should run randomized experiments in courses and curriculums to find what course content and structure is best for students.

Randomized experiments

Is it unethical to give different students different learning materials / resources?

Randomized experiments

Is it unethical to give different students different learning materials / resources?

Suppose you have 2 different R tutorials. You don’t know which one helps students learn R better. So you give each one to half the class. Two weeks later you check students’ R skills to decide which one to give out next year. Is that ok?

Randomized experiments

Medical research is based on randomized controlled trials that give different people different treatments.

Why should we not do the same in education?

Randomized experiments

Medical research is based on randomized controlled trials that give different people different treatments.

Why should we not do the same in education?

For learning, there is a false sense that we know what must be good. Often we are wrong about what works.

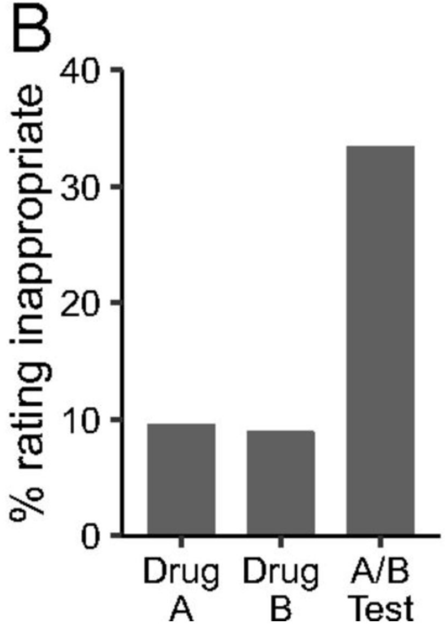
Evidence of Experiment Aversion

Objecting to experiments that compare two unobjectionable policies or treatments

Michelle N. Meyer, Patrick R. Heck, Geoffrey S. Holtzman, Stephen M. Anderson, William Cai, Duncan J. Watts, and Christopher F. Chabris

PNAS May 28, 2019 116 (22) 10723-10728; first published May 9, 2019 https://doi.org/10.1073/pnas.1820701116

There exist multiple FDA-approved blood pressure drugs and that, of these, “Doctor Jones” decides to prescribe all his patients a drug named simply “drug A” or “drug B” (in the A and B conditions), or decides to randomly assign his patients to receive drug A or drug B (in the A/B condition).



Major challenges ahead

Algorithmic bias in learning analytics

- We need to start measuring it, and reducing it

Life-long data governance

- What data do we store and how do we use it

Platform interoperability

- How can/should a student be tracked across different learning platforms