WEBINAR SERIES

A Conversation With ...

Michelle Seng Ah Lee

PhD Researcher, Compliant & Accountable Systems Research Group

University of Cambridge

Challenges of Algorithmic Fairness in Financial Services

CAIA ASSOCIATION®

FDP INSTITUTE™

The Global Designation for Finance Professionals in a Data-Driven Industry

The FDPi was created by CAIA to:

- ✓ Provide financial professionals with the knowledge necessary to succeed in an industry disrupted by the advent of big data and machine learning.
- ✓ Advocate for the highest levels of professional ethics and standards.
- ✓ Establish the FDP Charter as a global professional designation in the area of financial data science.

EARN YOUR FDP DESIGNATION

A globally-recognized charter is awarded to FDP charter holders

TWO ONLINE CLASSES*

Choose either Python or R

Can be completed before or after the exam

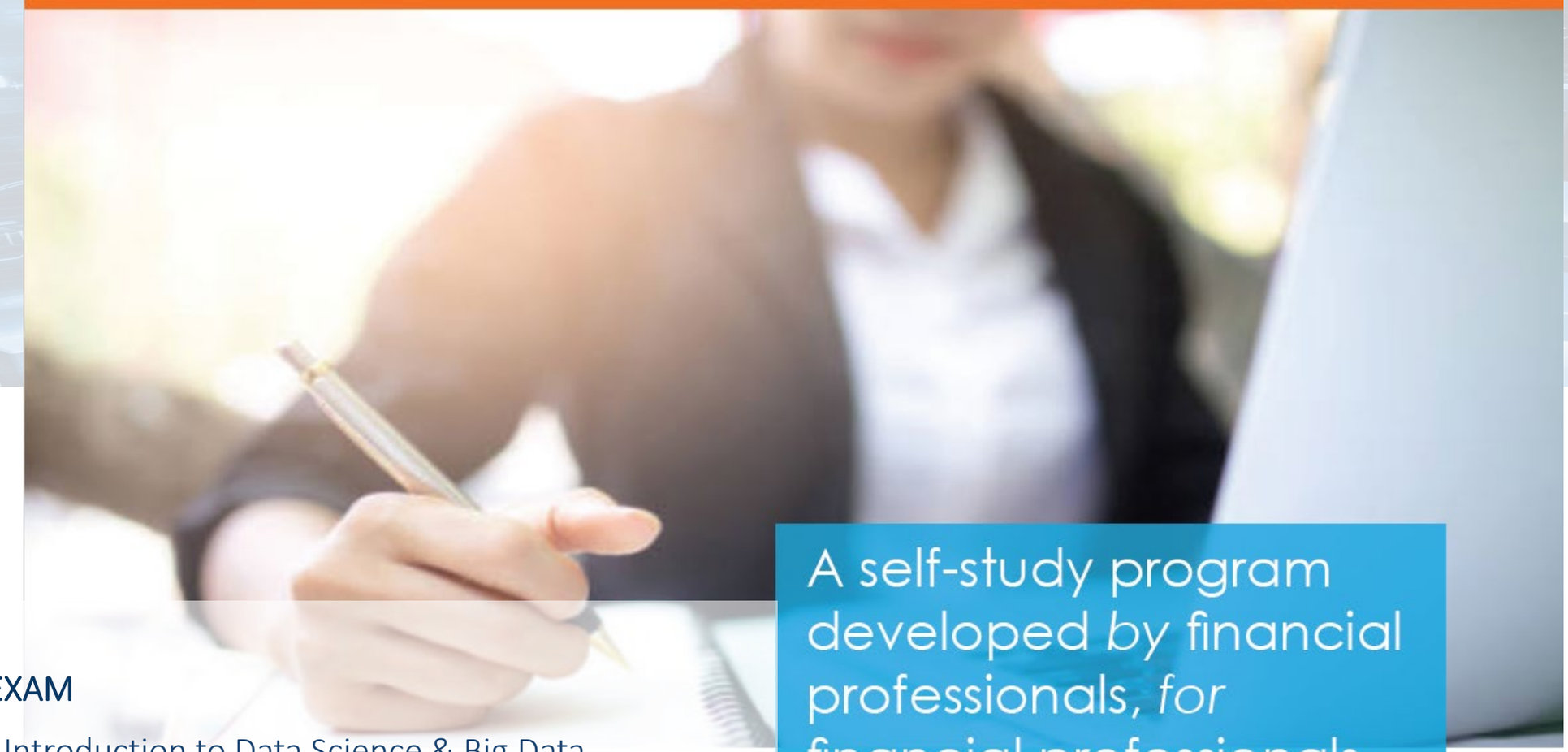
ONE COMPREHENSIVE EXAM

Offered twice per year

March & November

VALUE ADD

Employers increasingly seek to find professionals to have the skills to apply data science tools to solve their most challenging problems



A self-study program developed *by* financial professionals, *for* financial professionals.

FDP EXAM

1. Introduction to Data Science & Big Data
2. Machine Learning: Introduction to Algorithms
3. Machine Learning: Regression, Support Vector Machine & Time Series Models
4. Machine Learning: Regularization, Regression Trees, Random Forest & Overfitting
5. Machine Learning: Classification & Clustering
6. Machine Learning: Performance Evaluation, Backtesting & False Discoveries
7. Data Mining & Machine Learning: Naïve Bayes & Text Mining
8. **Big Data & Machine Learning: Ethical & Privacy Issues**
9. Big Data & Machine Learning in the Financial Industry

Exam candidates now **have one of two test options** for the FDP exam

Option 1: At a Prometric test center near you
Exam window: October 12 – November 8, 2020

Option 2: Through Remote proctoring from your home/office
Exam Day: December 1, 2020

Learn more at **www.fdpinstitute.org**

CAIA ASSOCIATION® FDP INSTITUTE™

The Global Designation for Finance Professionals in a Data-Driven INdustry

Financial Data Professional Institute

Challenges of Algorithmic Fairness in Financial Services

Why is it important, and what's next?

Michelle Seng Ah Lee
AI Ethics Lead, Deloitte UK
PhD Researcher,
University of Cambridge
Compliant & Accountable Systems Research Group

Keith Black
Managing Director,
Content Strategy,
CAIA

Mirjam Dekker
PM
FDP Institute

www.fdpinstitute.org

July 29, 2020

Todays webinar

- Challenges of Algorithmic Fairness in Financial Services
  - Why is it important, and what's next?

  - Introduction
  - The Problem Statement
  - Pre-AI: racial discrimination in lending (ex. US Mortgage)
  - Problem is not the algorithmic design
  - Why can't we just make algorithmics fair?
  - Fairness complexities
  - Fairness tests
  - Proposal: benchmarking
  - Trade-off between financial inclusion and minority denial rates
  - Why are algorithms biased against back applicants?
  - If algorithms can triangulate race, what data can we use?
  - Example: spelling error and default risk
  - Findings
  - Why do people make spelling errors?
  - Phd future research

CAIA ASSOCIATION®

FDP INSTITUTE™

UNIVERSITY OF CAMBRIDGE

Challenges of algorithmic fairness in financial services

Why is it important, and what's next?

Michelle Seng Ah Lee, PhD Candidate

About me

BA Political Science, Symbolic Systems
(Decision-making and rationality)

S&O Consulting,
Deloitte US

AI team,
Risk Analytics,
Deloitte UK

UK Chapter Lead, DataKind

MSc Oxford,
Digital Ethics Lab

Started PhD at
Cambridge

The problem statement

MO COMPARE Motorists fork out £1,000 more to insure their cars if their name is Mohammed

Top firms such as Admiral and Marks & Spencers have been dragged into an insurance race row after giving far lower quotes for drivers with traditionally English names like John

INVESTIGATION By Ben Leo
22nd January 2018, 12:17 am | Updated: 22nd January 2018, 11:51 pm

Harvard Business Review

ANALYTICS

Hiring Algorithms Are Not Neutral

by Gideon Mann and Cathy O'Neil

DECEMBER 09, 2016

SAVE SHARE 11 COMMENT TEXT SIZE PRINT $8.95 BUY COPIES

UK insurance company charges more based on email domain

Insurance firm Admiral is now the subject of an investigation by the UK Financial Conduct Authority based on its pricing policies.

Facial-Recognition Software Might Have a Racial Bias Problem

Depending on how algorithms are trained, they could be significantly more accurate when identifying white faces than African American ones.

CLARE GARVIE AND JONATHAN FRANKLE | APR 7, 2016 | TECHNOLOGY

The New York Times

Apple Card Investigated After Gender Discrimination Complaints

A prominent software developer said on Twitter that the credit card was "sexist" against women applying for credit.

E27 COMMUNITY FINTECH Igor Pesin

Your online, mobile, and social behaviour are now data-points used by fintech startups and governments in scoring credit-worthiness

Imagine a world where authoritarian governments and fintech companies mo everything you do, amass huge amounts of data on almost every interaction y and award you a single score that measures how "trustworthy" you are

Pre-AI: racial discrimination in lending (ex. US mortgage)

Data collection

Selection bias

Limited marketing / outreach to minority groups

Pre-application

Human bias

Black applicants quoted lower loan amounts, received less assistance and information (paired audit study)

Decision-making

Disparate impact

8% difference in denial rates even after controlling for all other “legitimate” features

Learning

Missing counterfactual

No information on whether or not those denied a loan would have defaulted

Problem is not the algorithm design

…and therefore cannot be solved algorithmically

Data collection

Selection bias

Requires a change in strategy to increase outreach and advertisement in high-minority areas

Pre-application

Human bias

Requires a change in process to provide similar service to similar individuals (vs. at individual discretion)

Decision-making

Disparate impact

Requires an understanding of **why** the existing credit risk process results in disadvantageous outcomes for minorities (e.g. income volatility) and to measure this difference

Learning

Missing counterfactual

Requires an active analysis of whether there are denials of credit-worthy applicants

Why can't we just make algorithms fair?

Definition	Philosophy	Contextual Limitations
Choose the model that makes everyone better off in aggregate: The most accurate model gives people what they "deserve" by minimising errors	*Expected consequentialism*	• Difference in distribution could be a reflection of inequality in other markets • Historical discrimination can inaccurately reflect outcomes • An algorithm is limited by the historical data available (selection bias)
Demographic parity, or group fairness: $P(\hat{Y} \mid A = 0) = P(\hat{Y} \mid A = 1)$.	*Strict egalitarianism*	• Ineffective where disproportionality in outcomes can be justified by non-protected, non-proxy attributes • Can lead to reverse discrimination and inaccurate predictions
Counterfactual fairness: $P(\hat{Y}_{A \leftarrow a}(U) = Y \mid X = x, A = a) = P(\hat{Y}_{A \leftarrow a'}(U) = y \mid X = x, A = a)$	*David Lewis, cause and effect*	• In most use cases, difficult or impossible to theorise a strong causal link and formalise its graphical representation
Individual fairness: $\hat{Y}(X^{(i)}, A^{(i)}) \approx Y((X^{(j)}, A^{(j)})$.	*Mixed*	• Difficult to define "similarity" that is independent of any protected attribute • Can embed bias in the similarity metric
Equalized odds: $P(\hat{Y} = 1 \mid A = 0, Y = y) = P(\hat{Y} = 1 \mid A = 1, Y = y), y \in \{0, 1\}$	*Dworkin's theory of Resource Egalitarianism*	• Fails to address discrimination that may already be embedded in the data • Market does not exist in a vacuum; impossible to isolate what is outside of people's control
Equalized opportunity:. $P(\hat{Y} = 1 \mid A = 0, Y = 1) = P(\hat{Y} = 1 \mid A = 1, Y = 1)$	*Dworkin's theory of Resource Egalitarianism*	• Fails to address discrimination that may already be embedded in the data • Focus on true positives may not be appropriate for some use cases

Fairness complexities

Each of these is presented as a one-size-fits-all precondition of fairness, but doesn't it depend on the context?

Fairness tests

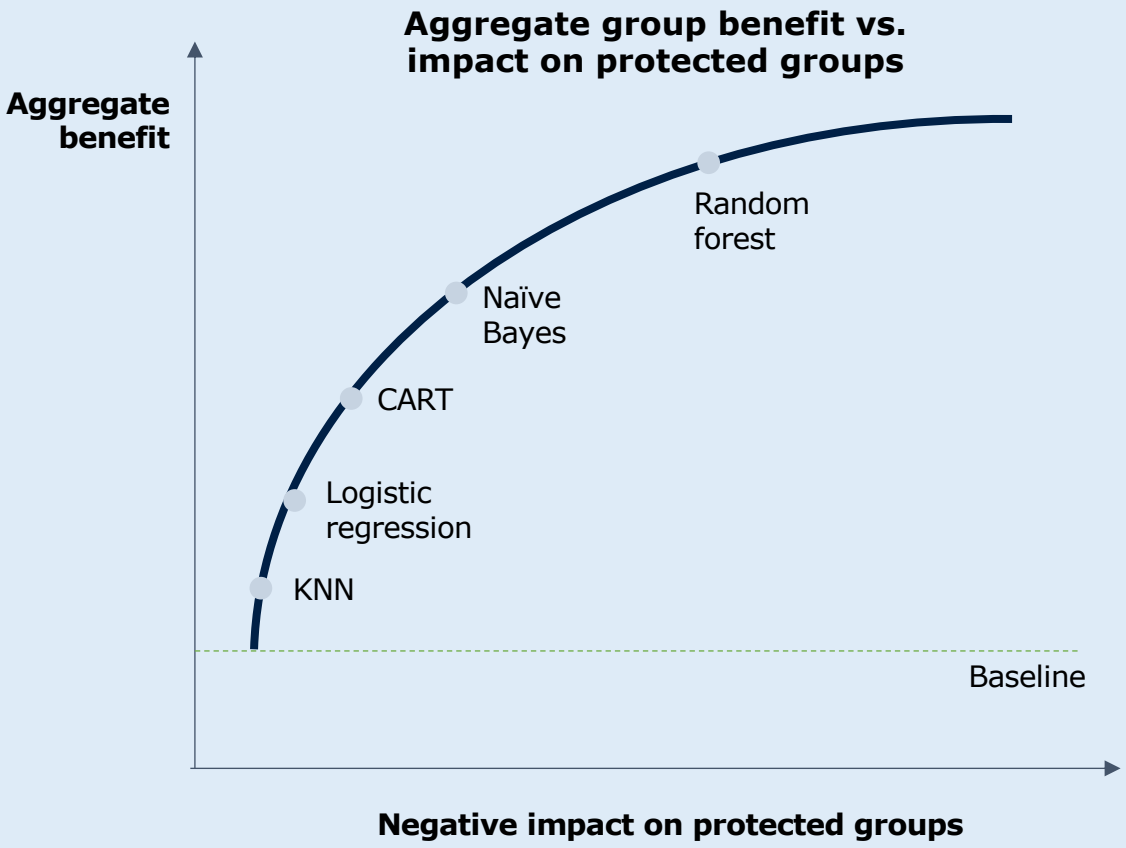
Approach in literature on fairness not useful in providing actionable insights

Model	Equality of Opportunity (EOP)	False Positive Error Rate Balance (FPERB)	Equal odds (EO)*	Positive Predictive Parity (PPP)	Positive Class Balance (PCB)	Negative Class Balance (NCB)
LR	4%	4%	4%	33%	2%	3%
KNN	6%	6%	6%	18%	8%	8%
CART	6%	6%	6%	21%	16%	18%
NB	7%	7%	7%	44%	41%	41%
RF	9%	9%	9%	3%	18%	18%

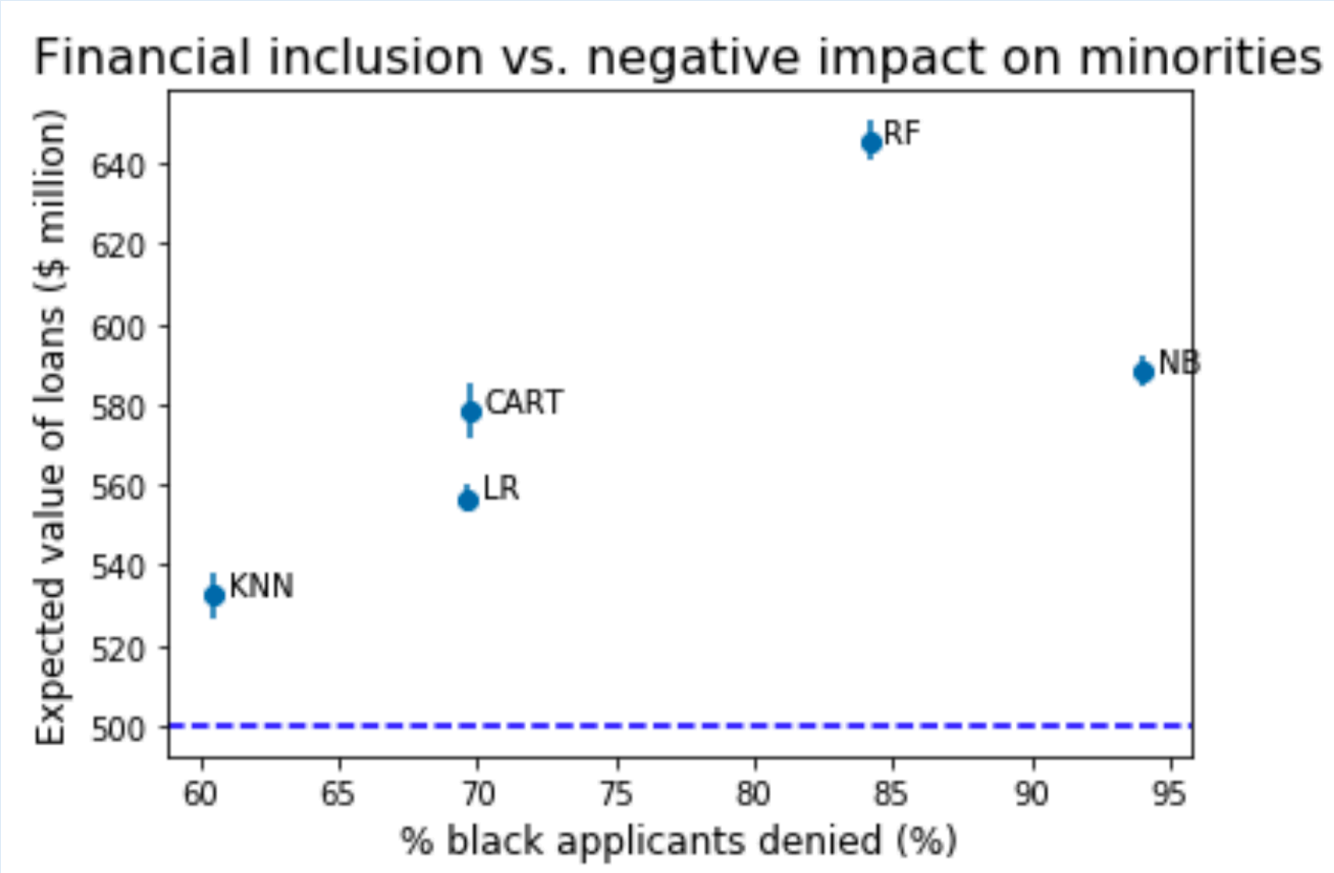
* EO condition is met if EOP and FPERB are both met.

My proposal: benchmarking

Which model's trade-off are you most comfortable with?



Trade-off between financial inclusion and minority denial rates



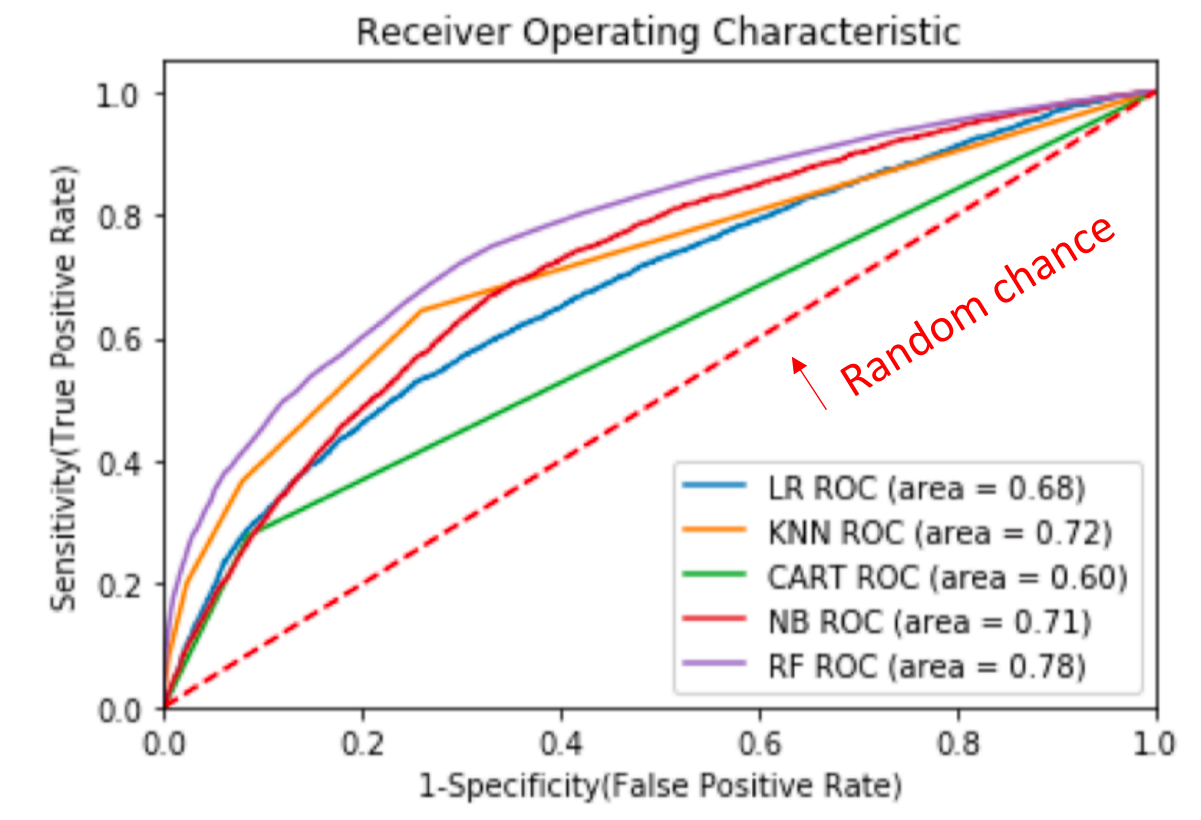
Why are the algorithms biased against black applicants?

Human bias

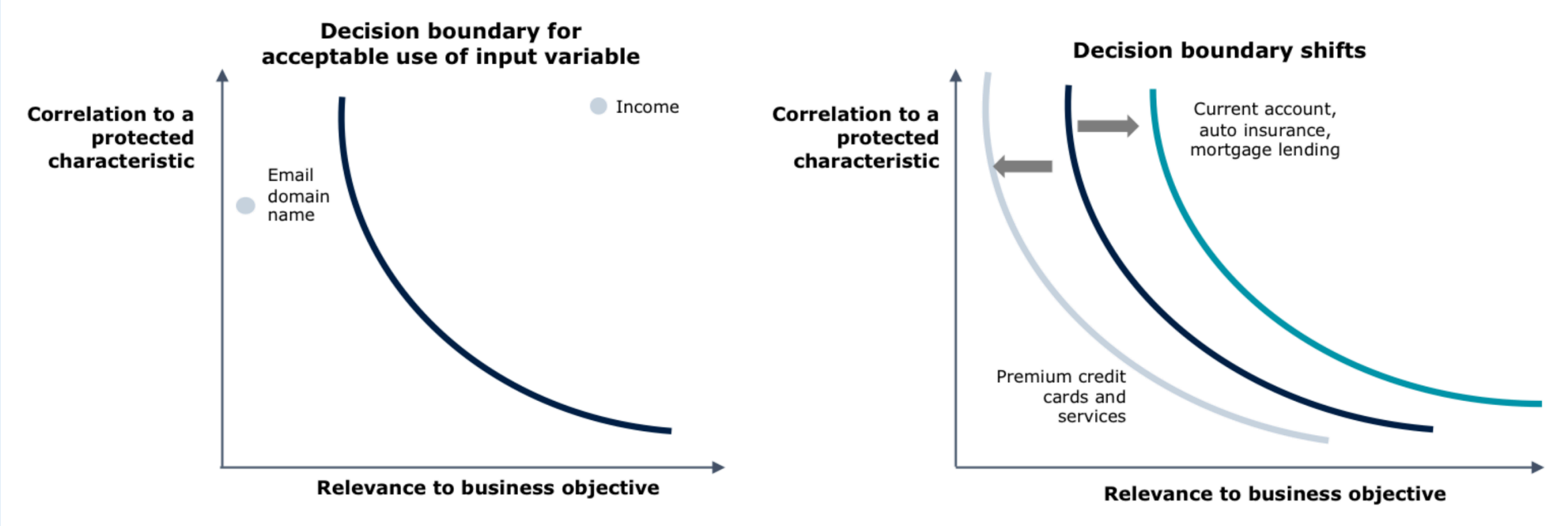
- FHA-insured loan type associated with both race and loan outcome

Embedded proxies

- It is possible to predict race using the input features



If algorithms can triangulate race, what data can we use?



Replicated from: https://sigai.acm.org/static/aimatters/5-2/AIMatters-5-2-07-Lee.pdf

Example: spelling error and default risk

- Open source peer-to-peer (P2P) lending data set
- Data from 2007-2011 retain the column containing free-form text description of the loan from the borrower (42,309 loans, 15% default rate)

Below is a partial excerpt from one of the descriptions, with non-standard language in *italics* and accidental typos in **bold**:

> Borrower added on 07/21/11 paying off house and 4,000.00 back to my saving *acct* bought a 12,000.00 boat and **payed** cash for it. Borrower added on 07/24/11 reason for paying house off i never have more then one loan at a time. Borrower added on 08/02/11 Thanks to the investors you're making a sure thing with me. my credit rating has **allways** been number one with me. thanks again

Findings

Table 6: Results: Logit on typos

Model:	Logit	No. Iterations:	6.0000
Dependent Variable:	loan status	Pseudo R-squared:	0.008
Date:	2019-04-20 01:12	AIC:	4871.2014
No. Observations:	5024	BIC:	4916.8553
Df Model:	6	Log-Likelihood:	-2428.6
Df Residuals:	5017	LL-Null:	-2447.8
Converged:	1.0000	Scale:	1.0000

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	-1.3897	0.1191	-11.6646	0.0000	-1.6232	-1.1562
Loan amount	0.0000	0.0000	3.2589	0.0011	0.0000	0.0000
Annual income	-0.0000	0.0000	-4.7294	0.0000	-0.0000	-0.0000
Number of words	-0.0000	0.0004	-0.0599	0.9522	-0.0007	0.0007
Total Levenshtein Distance	-0.0849	0.0604	-1.4062	0.1597	-0.2033	0.0334
% Phonetic Equivalent	-0.0294	0.0940	-0.3128	0.7544	-0.2136	0.1548
Number of Typos	0.1650	0.0771	2.1405	0.0323	0.0139	0.3160

Higher number of typos is associated with a greater probability of default.

P(default | 0 typo) = **17.35%**
P(default | 1 typo) = **19.84%**
P(default | 2 typos) = **22.59%.**

However, whether the typo was orthographic or phonological and the distance of the typo from the correction do not affect the likelihood of default.

Why do people make spelling errors?

Dyslexia

Non-native speaker

Carelessness

How do we isolate the impact of carelessness?

PhD future research questions

	Papers	Target journal / conference
Work in progress	A. From fairness metrics to Key Ethics Indicators (KEIs): a context-conscious approach to algorithmic ethics in an unequal society	*Science and Engineering Ethics Journal*
	B. Identifying 6 types of biases in software development lifecycle	*Fairness, Accountability, Transparency (FAccT) Conference*
	C. Gaps in the open source fairness and explanation toolkits	*ACM CHI Conference*
	D. Regulating beyond AI and ADM: risk factor framework	*European Data Protection Handbook* (proposal accepted as a chapter for publication in 2021)

Works cited / references

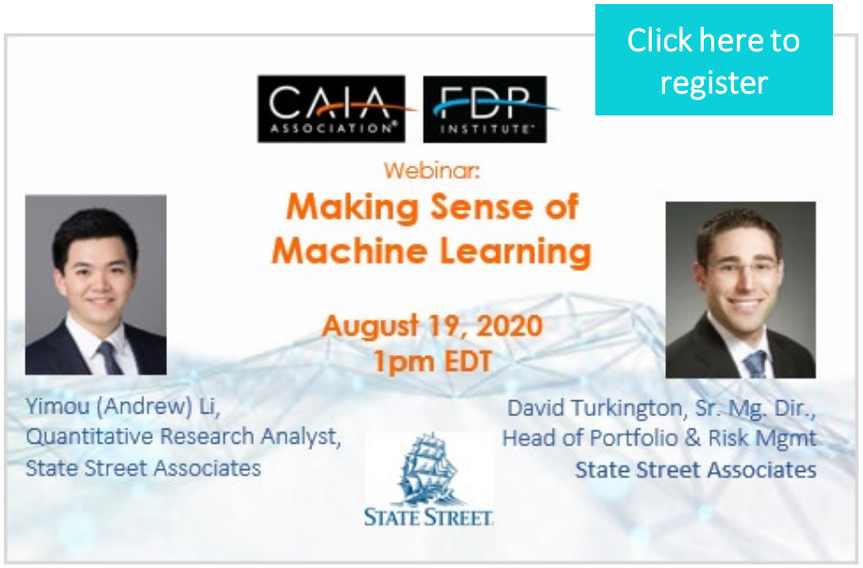
- List of my publications: https://michellesengahlee.wordpress.com/publications/
- My blog posts on Deloitte Financial Services: https://ukfinancialservicesinsights.deloitte.com/u/102fti4/michelle-lee

Recommended further reading:

- 50 years of (test) unfairness: https://arxiv.org/abs/1811.10104
- Delayed impact of fair machine learning: https://arxiv.org/abs/1803.04383
- The measure and mismeasure of fairness: https://arxiv.org/abs/1808.00023
- Inherent trade-offs in fair determination of risk scores: https://arxiv.org/abs/1609.05807
- 21 definitions of fairness and their politics: https://www.youtube.com/watch?v=jIXIuYdnyyk

Q & A

Add your questions to the chat room



CAIA ASSOCIATION® FDP INSTITUTE™

In Closing

Learn more about the FDP Institute at www.fdpinstitute.org