Machine Learning and Risk Modelling: Challenges and Advances

Ali Ashtari

Scotiabank

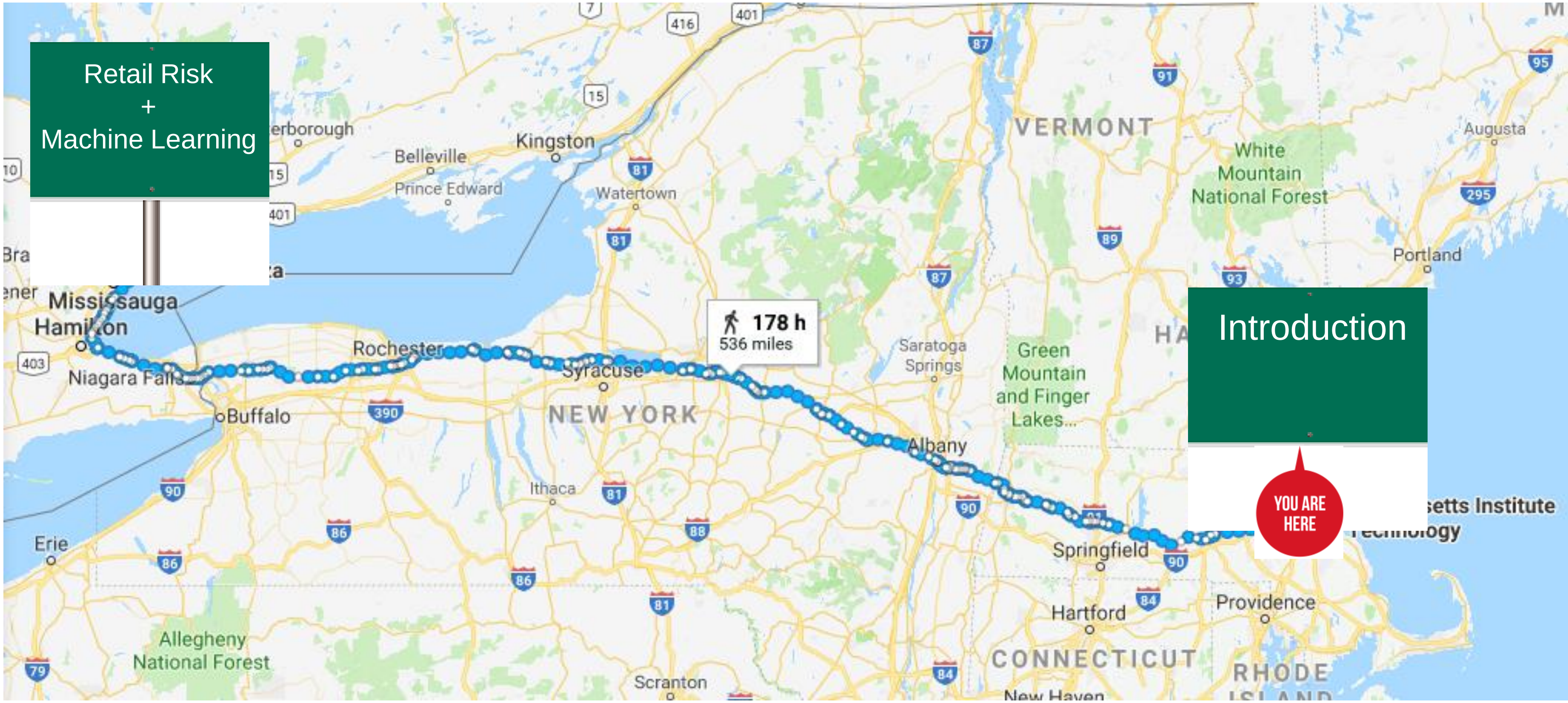
2 April 2018

Disclaimer

CHATHAM HOUSE
The Royal Institute of International Affairs

- The views expressed in this presentation are the personal views of the speaker and do not necessarily reflect the views or policies of current or previous employers.
- Participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed.

Our journey for the next 45 minutes

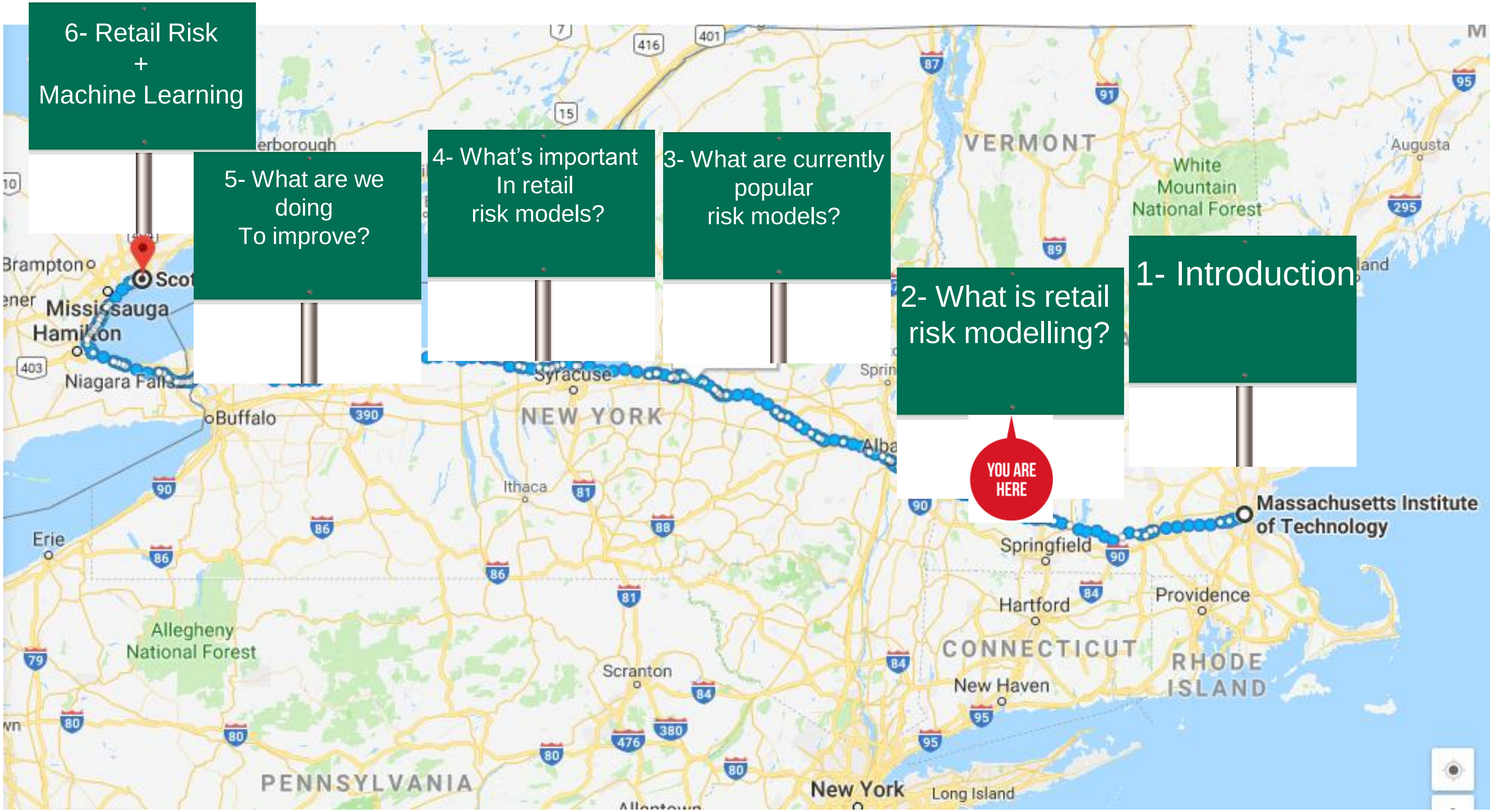


John and Joe will join us in the journey

John
A data scientist, also a gangster

Joe
A teacher, also a nice person

Where are we?



6- Retail Risk + Machine Learning
5- What are we doing To improve?
4- What's important In retail risk models?
3- What are currently popular risk models?
2- What is retail risk modelling?
YOU ARE HERE
1- Introduction
VERMONT
NEW YORK
PENNSYLVANIA
CONNECTICUT
RHODE ISLAND
Massachusetts Institute of Technology
Mississauga
Hamilton
Niagara Falls
Buffalo
Syracuse
Ithaca
Springfield
Hartford
Providence
New Haven
New York
Long Island
Scranton
Erie
Brampton
Augusta
White Mountain National Forest
Allegheny National Forest

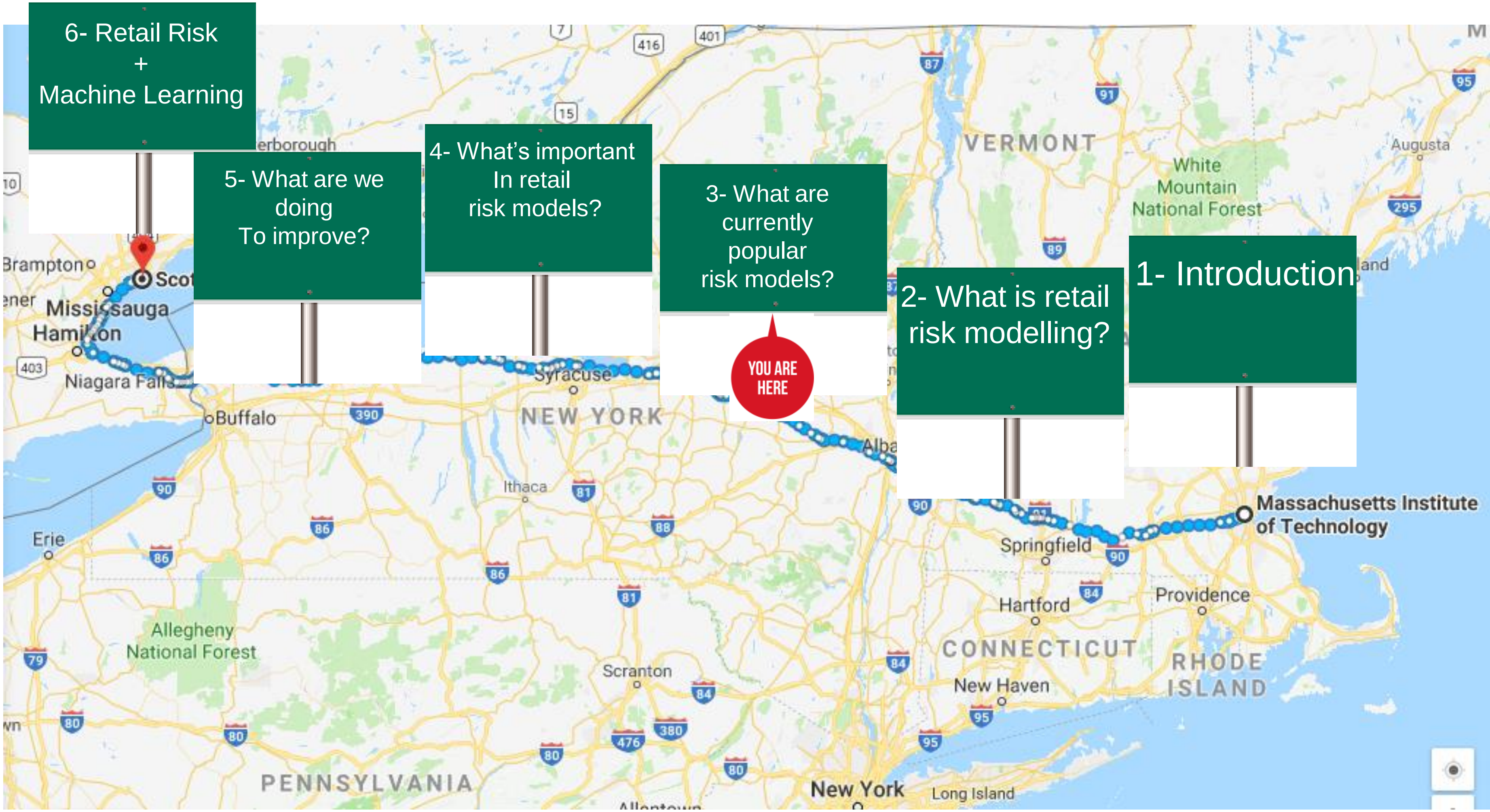
Retail Credit Risk Modelling

- Products
  - Credit cards
  - Autoloan
  - Mortgages
  - …
- Events
  - Delinquency
  - Default
  - Bankruptcy
- Stages
  - Origination
  - Account management
  - Collection

"We are prepared to make you a loan, but first you have to prove that you really don't need it."

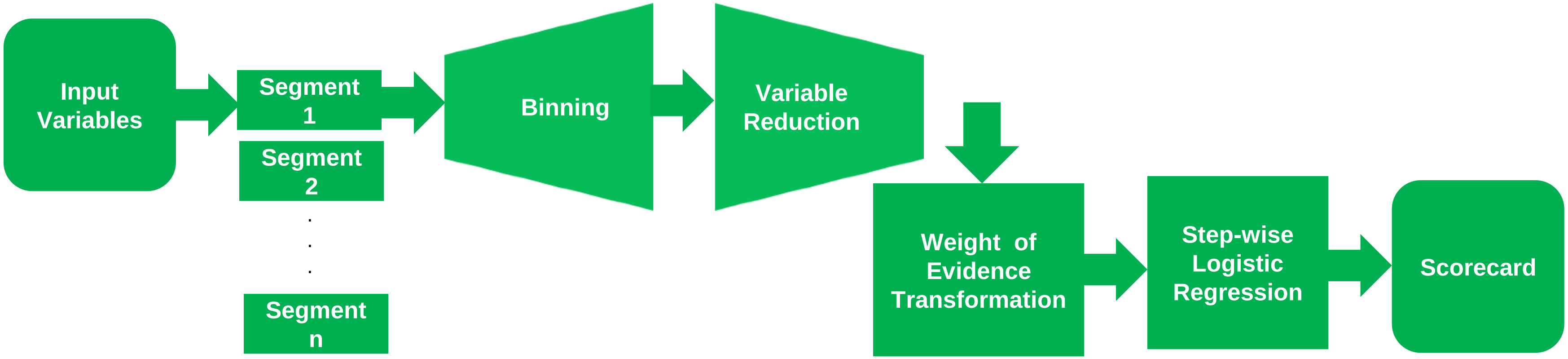
We will see several examples of how bad retail risk models could lead a bank to support the gangster and screw the nice person

Where are we?


6- Retail Risk
+
Machine Learning
5- What are we doing
To improve?
4- What's important
In retail
risk models?
3- What are currently popular
risk models?
YOU ARE HERE
2- What is retail risk modelling?
1- Introduction
VERMONT
NEW YORK
PENNSYLVANIA
CONNECTICUT
RHODE ISLAND
Massachusetts Institute of Technology
Brampton
Mississauga
Hamilton
Niagara Falls
Buffalo
Syracuse
Ithaca
Erie
Allegheny National Forest
Scranton
New York
Long Island
New Haven
Hartford
Springfield
Providence
White Mountain National Forest
Augusta

“Scorecards”

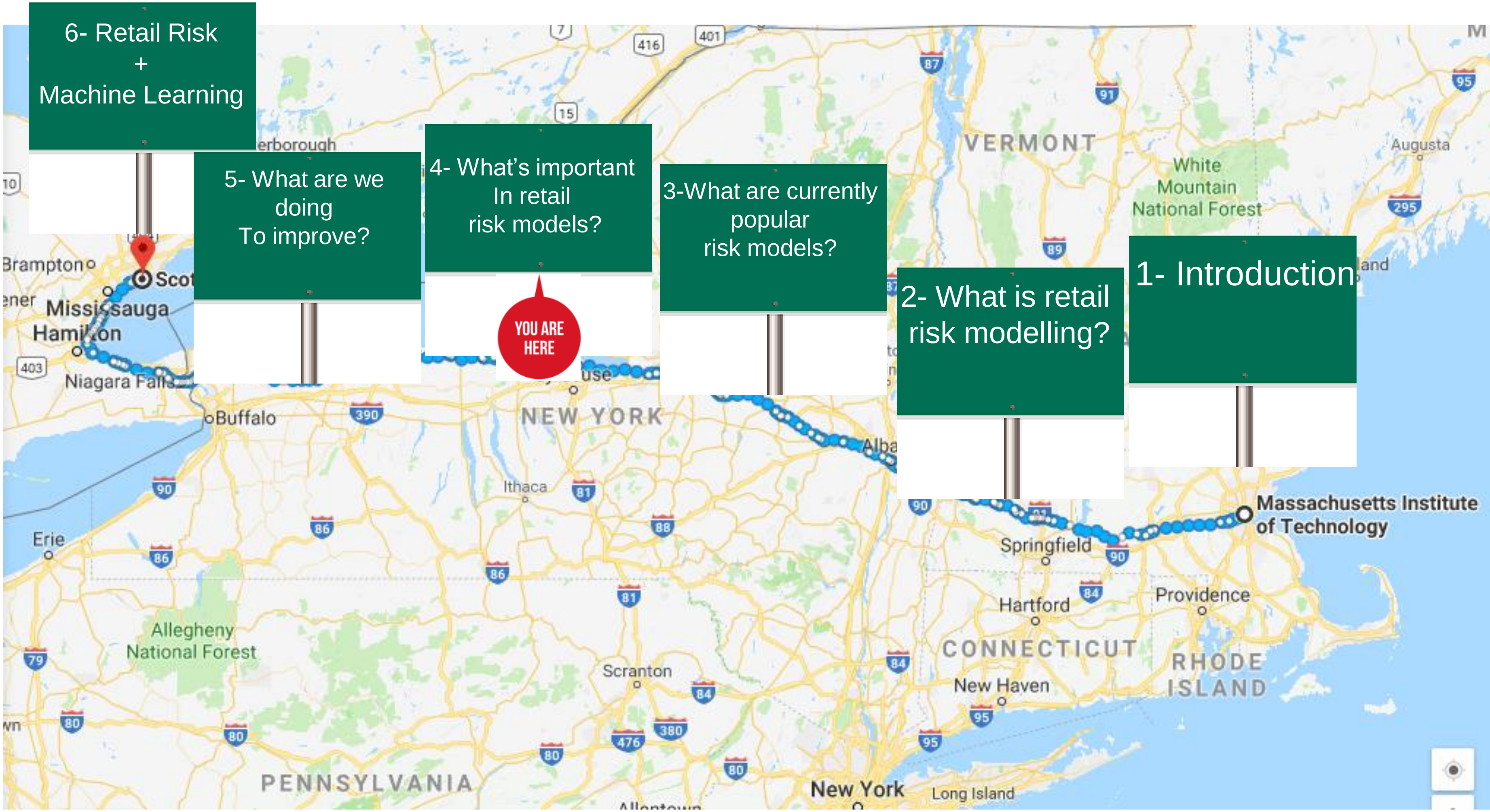
- Segmented Logistic Regression on Weight of Evidence of variables binned and reduced.



  - Predictive performance: Kind of OK
  - Interpretability: OK
  - Robustness: Kind of OK

- **Most importantly, it’s easy to understand.**

Where are we?


6- Retail Risk
+
Machine Learning
5- What are we doing To improve?
4- What's important In retail risk models?
YOU ARE HERE
3-What are currently popular risk models?
2- What is retail risk modelling?
1- Introduction
Massachusetts Institute of Technology
VERMONT
NEW YORK
CONNECTICUT
RHODE ISLAND
PENNSYLVANIA
Mississauga
Hamilton
Brampton
Niagara Falls
Buffalo
Erie
Ithaca
Scranton
Springfield
Hartford
Providence
New Haven
New York
Long Island
Augusta
White Mountain National Forest
Allegheny National Forest

What do we care about in retail risk models?

1. Predictive performance
2. Interpretability
3. Robustness

Typical machine learning applications care most about predictive performance whereas in retail risk models these three factors are almost equally important. Therefore:

IT'S

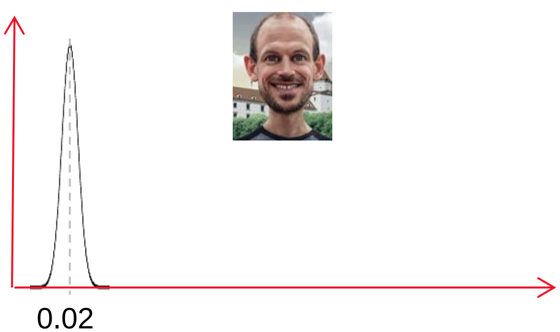
COMPLICATED

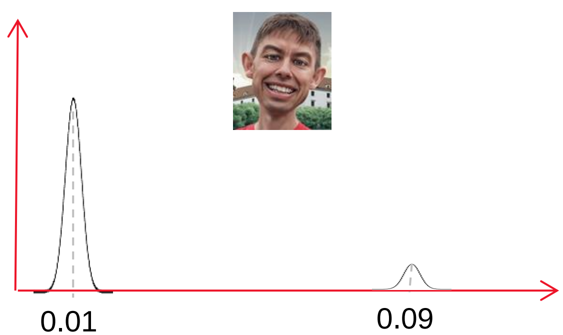
What do we care about in retail risk models?

1. Predictive performance (2 slides)
2. Interpretability (2 slides)
3. Robustness (2 slides)

Predictive performance

- Performance of what?
  - Model error vs. business impact (risk adjusted margin)
    - $r(\alpha) = \int u(Y, \alpha) \mathrm{d}P(Y|\alpha)$
- Consider a money laundry example:
- $u = \begin{cases} 1 & \mathit{if\ not\ money\ laundry} \\ -999 \ldots 9 & \mathit{if\ money\ laundry} \end{cases}$

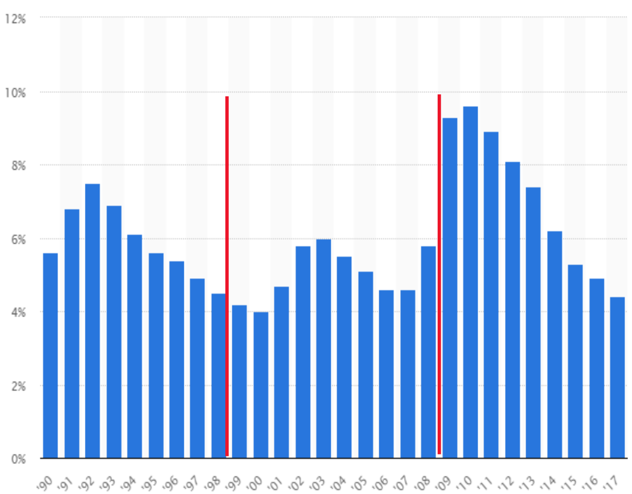




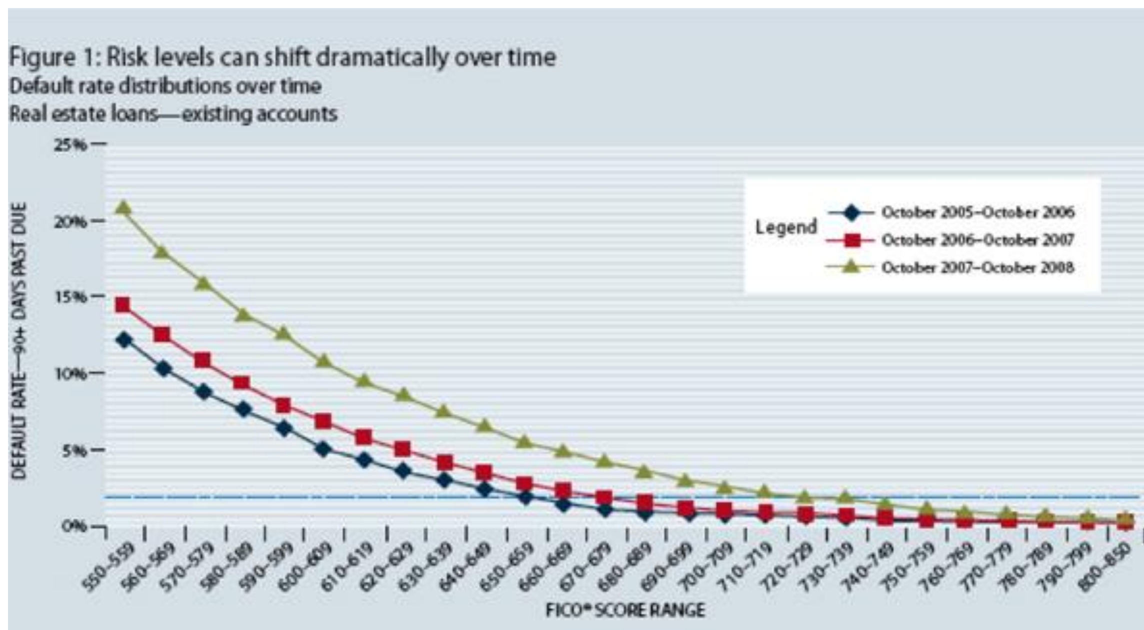
Predictive performance

- Non-stationarity
  - Unemployment rate, key interest rates, GDP, housing prices … are non-stationary.
  - Their impact on retail risk is significant.

Unemployment rate 1990-2017



statista.com



(Jennings, 2017)

What do we care about in retail risk models?

1. Predictive performance (2 slides)
2. Interpretability (2 slides)
3. Robustness (2 slides)

Interpretability

- Fairness requires interpretability
  - No discrimination based on race, color, religion, national origin, sex, marital status, age, source of income …
  - Example:  claims we did not give him a credit card because of his race.

Joy Buolamwini, MIT

Algorithmic Bias : Automated Facial Analysis

MIT Media Lab

Advised By: Ethan Zuckerman, Mitch Resnick, Hal Abelson

Machine learning is increasingly part of daily life. We have learned how to use data sets to teach machines to detect faces and predict patterns. However bias in training data leads to bias in the systems that are created. Algorithmic bias leads to exclusionary experiences and practices. Joy is establishing tools and methods for full spectrum testing of computer vision libraries to mitigate algorithmic bias that leads to poor detection of faces that are not well represented in current training sets.


Support The Guardian | Subscribe | Find a job | Sign in | Search | International edition

The Guardian

News | Opinion | Sport | Culture | Lifestyle | More

Money ▸ Property Pensions Savings Borrowing Careers

Inequality

Rise of the racist robots – how AI is learning all our worst impulses

There is a saying in computer science: garbage in, garbage out. When we feed machines data that reflects our prejudices, they mimic them – from antisemitic chatbots to racially biased software. Does a horrifying future await people forced to live at the mercy of algorithms?

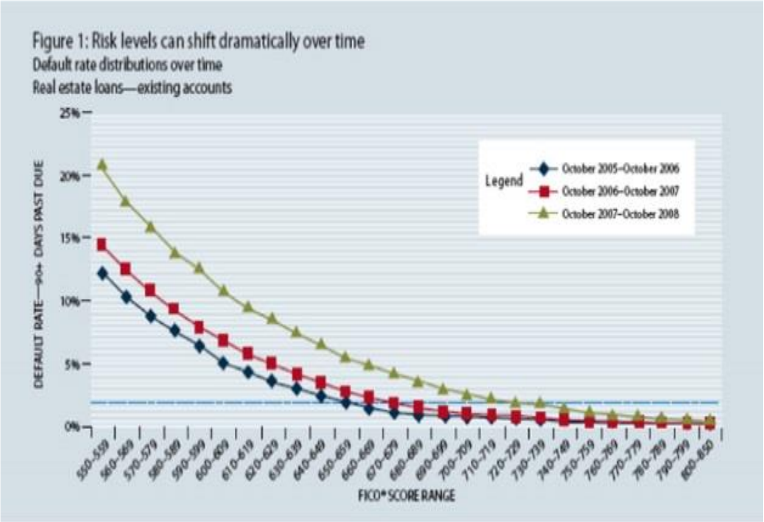
Advertisement

Interpretability

- Compensating model shortcomings by using domain knowledge requires interpretability
- Two typical shortcomings:
  - Point estimates instead of distributions

$$r(\alpha) = \int u(Y, \alpha) \mathrm{d}P(Y|\alpha)$$

  - Assuming stationarity



(Jennings, 2017)

What do we care about in retail risk models?

1. Predictive performance (2 slides)
2. Interpretability (2 slides)
3. Robustness (2 slides)

Robustness

Adversarial

(Evtimov, 2018)

Risk modelling in the industry usually uses inputs that are aggregated over time and thus are hard to perturb → More resilient against adversarial attacks.

Operational

thoughtco.com

If a datafeed goes down, will the model produce significantly different results?

The models use few predictive variables and thus are less operationally resilient. The industry deals with this issue through monitoring and reporting.

Robustness, a bit of formalism.

- Consider models M_1 and M_2, event E, utility U, data D and perturbed data $D(\mathrm{E})$
  - E is either a local (triggered by an individual), group (triggered by a group of individuals), or global event.
  - E can have an observable effect on data either at the time of training or at the time of inference/prediction. This effect perturbs data D during training/inference to $D(\mathrm{E})$.
  - $M_{1,2}(D) \neq M_{1,2}\big(D(E)\big)$
- M_1 is more robust than M_2 (modulo E, U) if:
  - $U(M_1, D(E)) > U(M_2, D(E))$
- This definition covers both adversarial and operational robustness.
- Scotia is funding research on robustness at MIT through Systems That Learn.

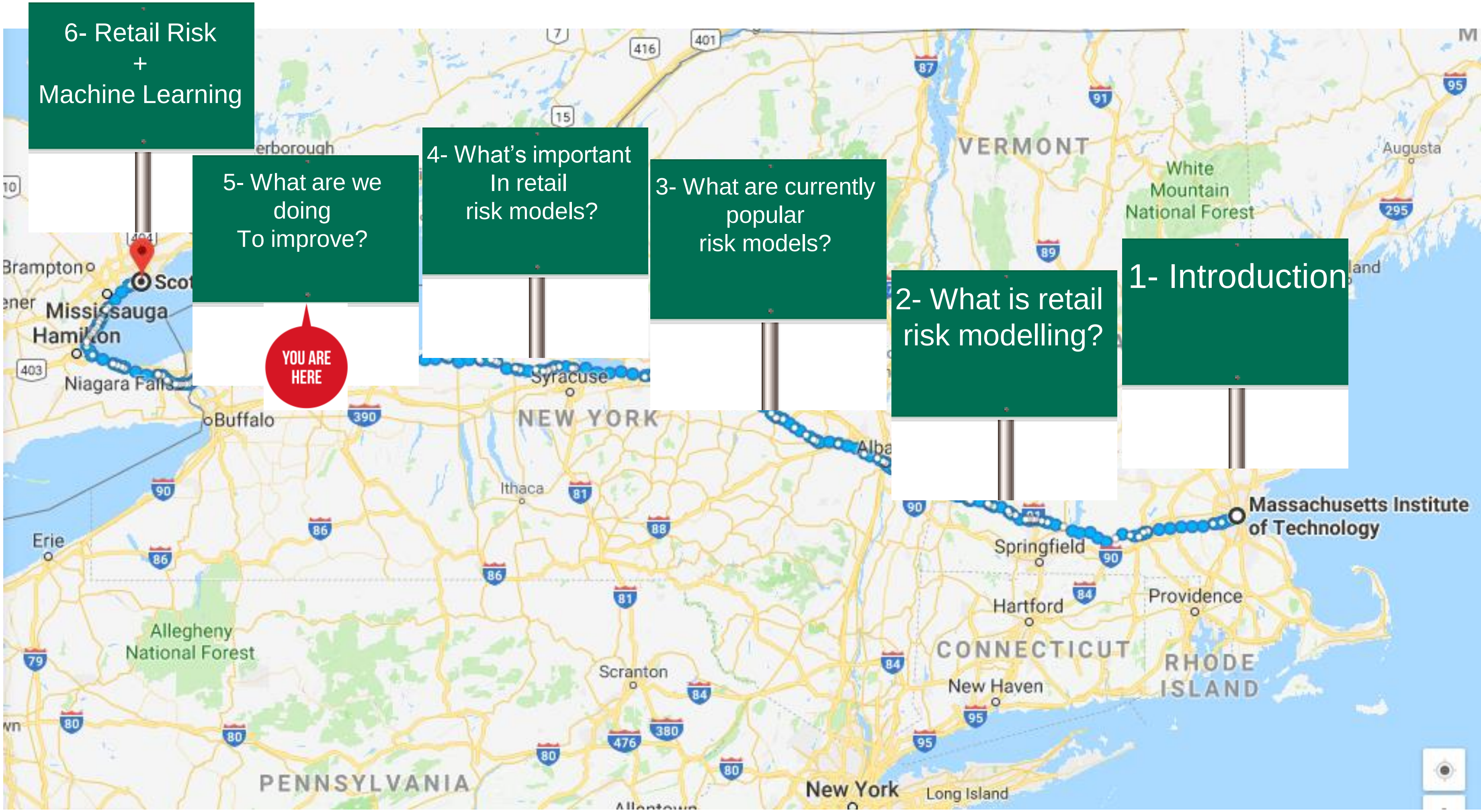
SYSTEMS THAT LEARN @ CSAIL

Distributionally Robust Deep Learning as a Generalization of Adversarial Training

Matthew Staib
MIT CSAIL
mstaib@mit.edu

Stefanie Jegelka
MIT CSAIL
stefje@mit.edu

Where are we?



6- Retail Risk
+
Machine Learning
5- What are we
doing
To improve?
YOU ARE
HERE
4- What's important
In retail
risk models?
3- What are currently
popular
risk models?
2- What is retail
risk modelling?
1- Introduction
Massachusetts Institute
of Technology
Mississauga
Hamilton
Brampton
Niagara Falls
Buffalo
Syracuse
NEW YORK
Ithaca
VERMONT
White
Mountain
National Forest
Augusta
Springfield
Hartford
Providence
CONNECTICUT
RHODE
ISLAND
New Haven
New York
Long Island
Scranton
Erie
Allegheny
National Forest
PENNSYLVANIA

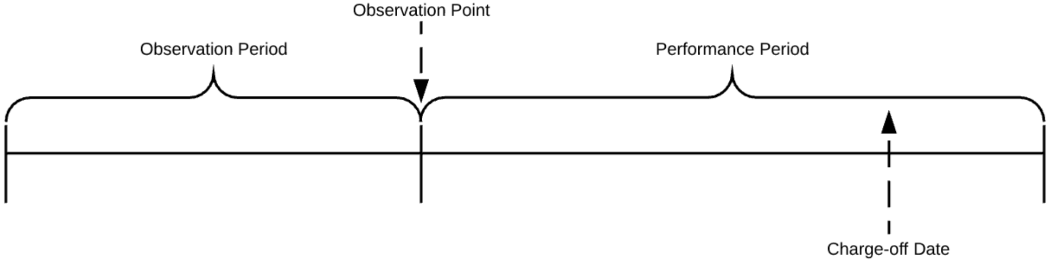
How are we improving?

- We are doing lots of stuff!
  - Deep learning
  - Interpreting “black box” models
  - Bayesian approaches
  - Representation learning

Deep learning project with Deep™ Learni.ng – deployed

- Credit card collections: Which delinquent customers to email or call?

- **BIG DATA!**
  - Millions of applications
  - Thousands of variables
  - Decades of history



THE WALL STREET JOURNAL.

Subscribe

Home World U.S. Politics Economy Business Tech Markets Opinion Life & Arts Real Estate WSJ. Magazine

CIO JOURNAL.

Scotiabank Deploys Deep Learning to Improve Credit Card Collections

Tool helps identify potentially delinquent customers and suggests ways to approach them

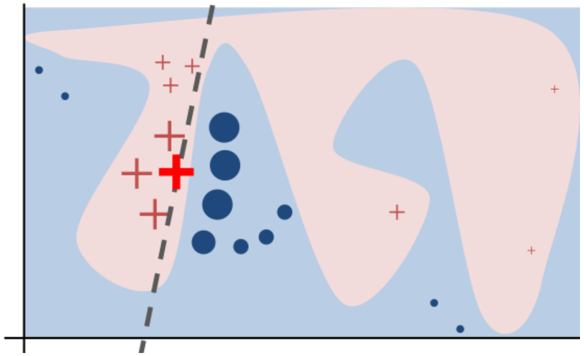
By Steven Norton

0 COMMENTS

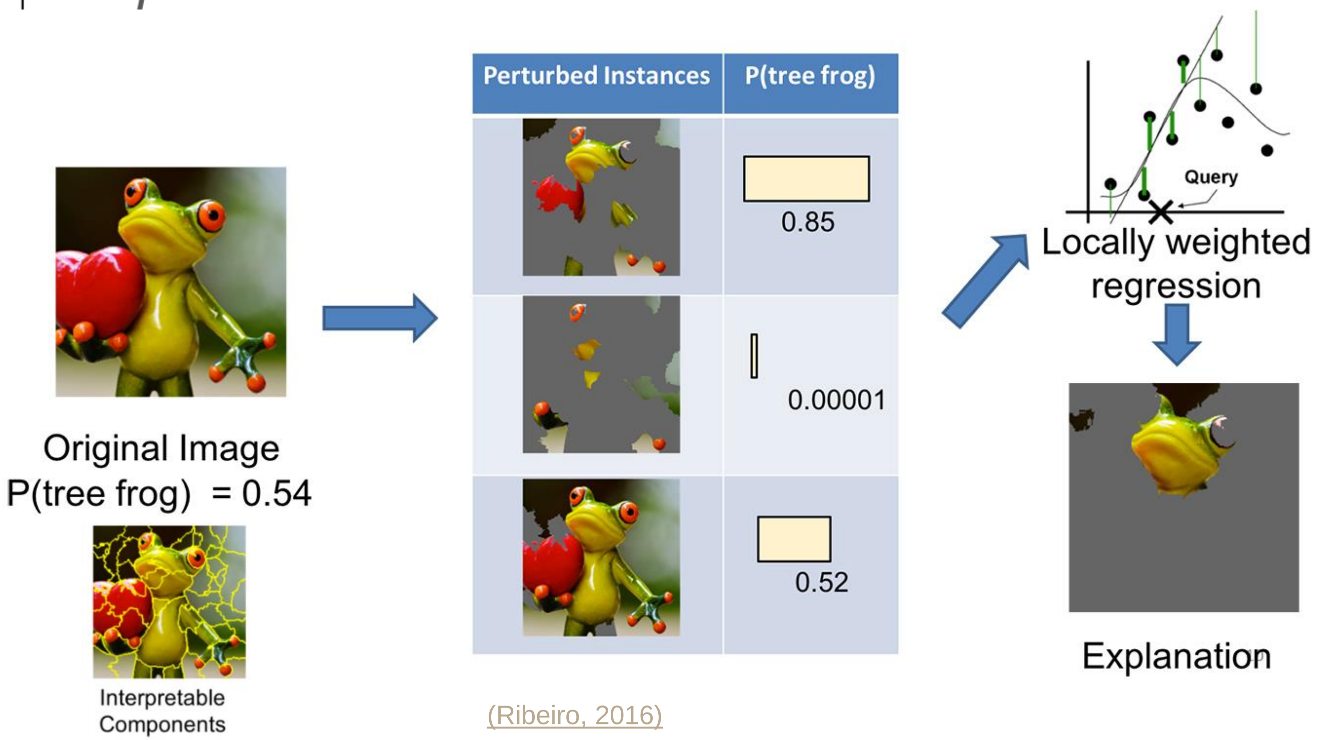
Feb 6, 2017 5:35 pm ET

Interpreting Deep Learning with LIME

Local Interpretable Model-agnostic Explanations



(a) Husky classified as wolf (b) Explanation



(Ribeiro, 2016)

That slide that only had equations.

let G be the class of linear models, such that $g(z') = w_g \cdot z'$

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \;\; \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

$$\pi_x(z) = exp(-D(x, z)^2/\sigma^2)$$

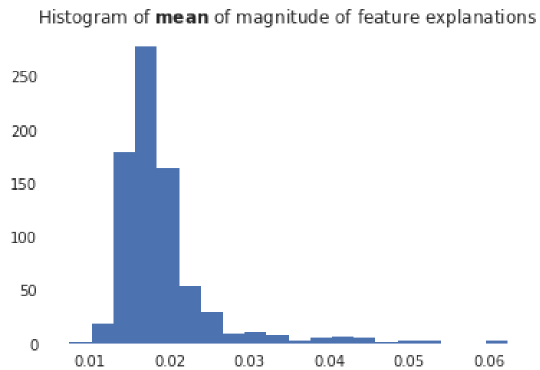
$$\mathcal{L}(f, g, \pi_x) = \sum_{z, z' \in \mathcal{Z}} \pi_x(z) \left(f(z) - g(z')\right)^2$$

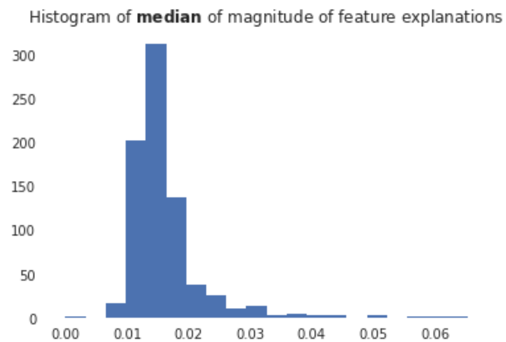
Require: Classifier f, Number of samples N
Require: Instance x, and its interpretable version x'
Require: Similarity kernel π_x, Length of explanation K
  $\mathcal{Z} \leftarrow \{\}$
  for $i \in \{1, 2, 3, ..., N\}$ **do**
    $z'_i \leftarrow sample\_around(x')$
    $\mathcal{Z} \leftarrow \mathcal{Z} \cup \langle z'_i, f(z_i), \pi_x(z_i) \rangle$
  end for
  $w \leftarrow$ K-Lasso$(\mathcal{Z}, K)$ $\triangleright$ with z'_i as features, $f(z)$ as target
  return w

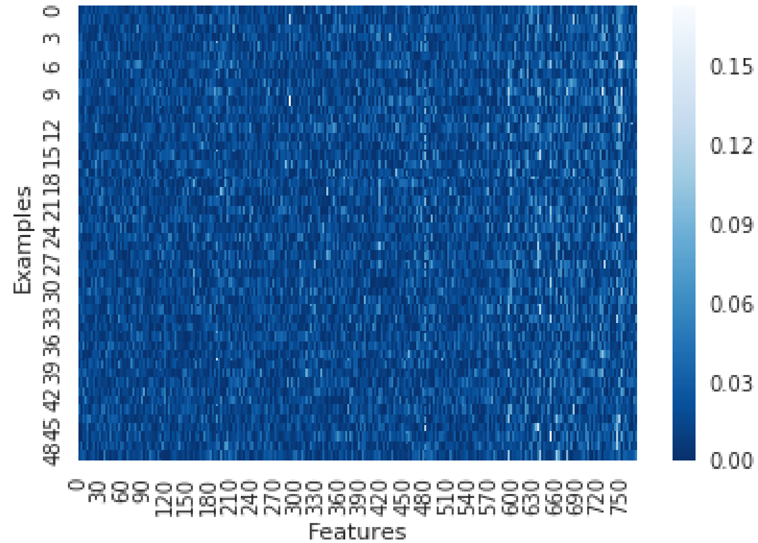
(Ribeiro, 2016)

LIME Results on our Deep learning model

- Can interpret every decision!
- We are considering new research in the area of interpretability

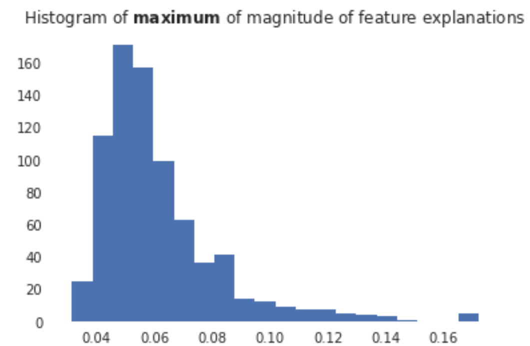






A causal framework for explaining the predictions of black-box sequence-to-sequence models

David Alvarez-Melis and **Tommi S. Jaakkola**
CSAIL, MIT



Building Machines That Learn and Think Like People

Brenden M. Lake,[1] Tomer D. Ullman,[2,4] Joshua B. Tenenbaum,[2,4] and Samuel J. Gershman[3,4]
[1]Center for Data Science, New York University
[2]Department of Brain and Cognitive Sciences, MIT
[3]Department of Psychology and Center for Brain Science, Harvard Universi
[4]Center for Brains Minds and Machines

Rationalizing Neural Predictions

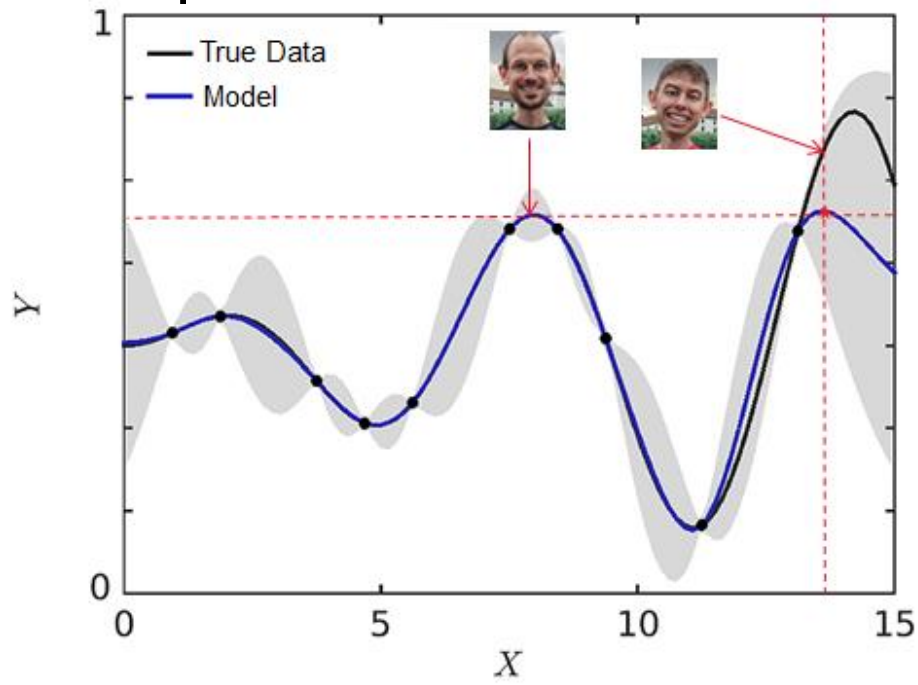
Tao Lei, Regina Barzilay and Tommi Jaakkola
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
{taolei, regina, tommi}@csail.mit.edu

Bayesian approaches

- Bayesian approaches give us probability distributions as opposed to point estimates. <u>This is VERY important.</u>

$$r(\alpha) = \int u(Y, \alpha)\mathrm{d}P(Y|\alpha)$$

- Use confidence in predictions for decision making



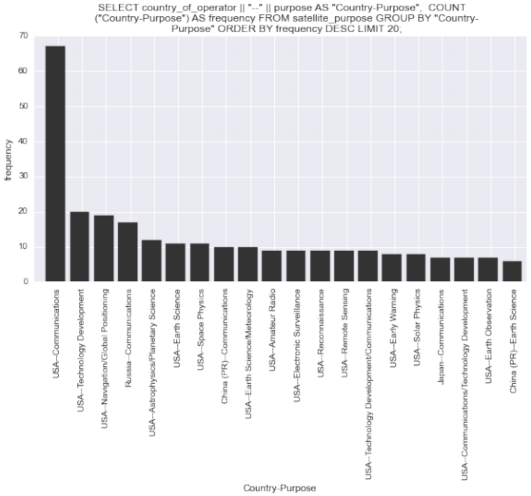
Bayesian approaches

- BayesDB (Vikash Mansinghka: MIT)
- Query the probable implications of data
- Example: The most probable countries and purposes of a satellite with a 500 kilogram dry mass in geosynchronous orbit
- `SELECT country_of_operator, purpose, Class_of_orbit, Dry_mass_kg FROM satellites WHERE Class_of_orbit = "GEO" AND Dry_Mass_kg BETWEEN 400 AND 600;`

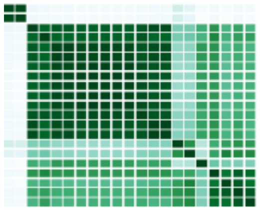
	Country_of_Operator	Purpose	Class_of_Orbit	Dry_Mass_kg
0	India	Communications	GEO	559
1	India	Meteorology	GEO	500

- `SIMULATE country_of_operator, purpose FROM satellites GIVEN Class_of_orbit = GEO, Dry_mass_kg = 500 LIMIT 1000;`

- It’s a US communication satellite! →
- Another usage of bayesDB: probability of dependence



Correlation

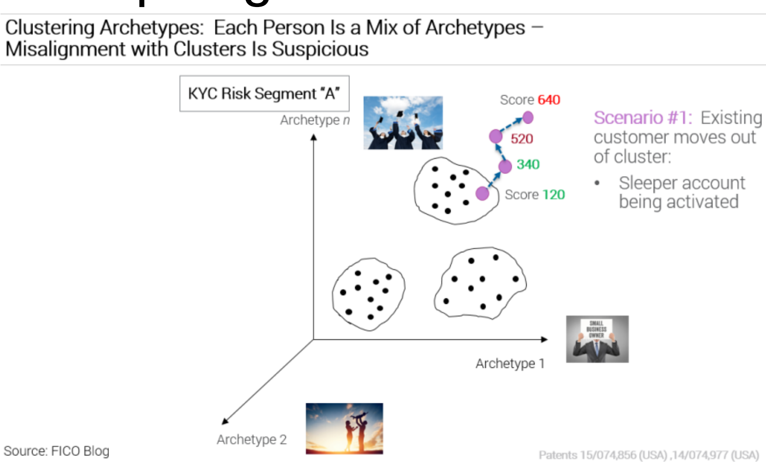


Probability of dependence

(Mansinghka, 2015)

Anti money laundry models

- Challenges:
  - Very few alarms in huge datasets
  - From those few alarms, most are false
  - Few confirmations for true alarms
- Unsupervised approaches:
  - Distance based
  - Learning underlying structure
  - Topological



Fico.com

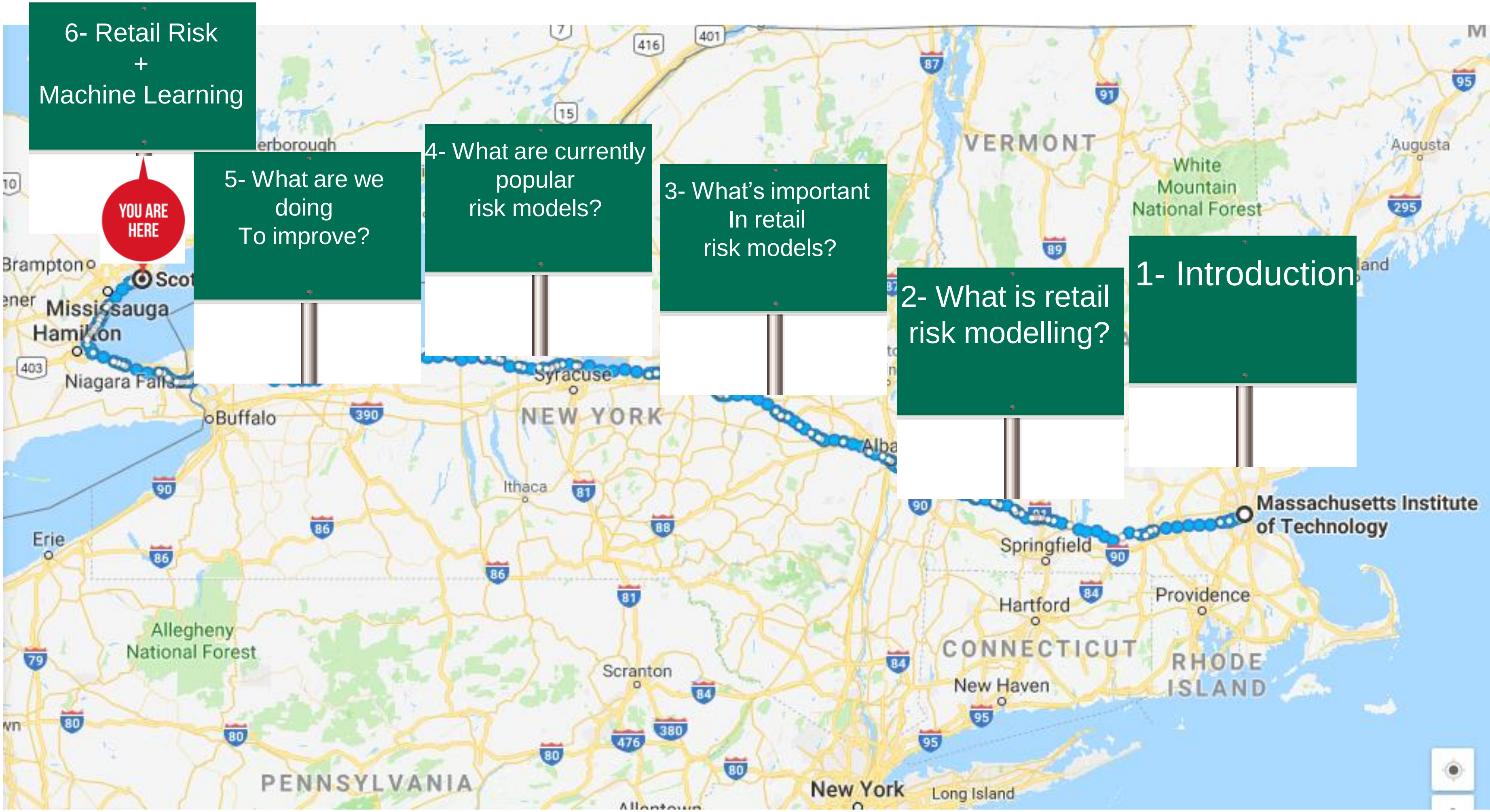

REUTERS World Business Markets Politics TV

Myanmar | CERAWeek | Brexit | North Korea | Earnings | Future of Money | Breakingviews

BUSINESS NEWS JUNE 1, 2017 / 6:07 AM / 9 MONTHS AGO

HSBC partners with AI startup to combat money laundering

Where are we?


6- Retail Risk
+
Machine Learning
YOU ARE HERE
5- What are we doing
To improve?
4- What are currently popular
risk models?
3- What's important
In retail
risk models?
2- What is retail risk modelling?
1- Introduction
Massachusetts Institute of Technology
VERMONT
NEW YORK
CONNECTICUT
RHODE ISLAND
PENNSYLVANIA
White Mountain National Forest
Allegheny National Forest
Augusta
Brampton
Mississauga
Hamilton
Niagara Falls
Buffalo
Syracuse
Ithaca
Erie
Scranton
Springfield
Hartford
Providence
New Haven
New York
Long Island

References

- Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2016.

- Ivan Evtimov, Kevin Eykholt, Earlence Fernandes, Tadayoshi Kohno, Bo Li, Atul Prakash, Amir Rahmati, Dawn Song, Robust Physical-World Attacks on Deep Learning Visual Classification, Computer Vision and Pattern Recognition CVPR, June 2018.

- Feras Saad, ikash Mansinghka, A Probabilistic Programming Approach To Probabilistic Data Analysis, Advances in Neural Information Processing Systems (NIPS), December 2016.

- Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, Samuel J. Gershman, Building machines that learn and think like people, Behavioral and Brain Sciences, Volume 40, 2017.

- Tao Lei, Regina Barzilay, Tommi Jaakkola, Rationalizing Neural Predictions, Proceedings of Conference on Empirical Methods in Natural Language Processing, November 2016.

- David Alvarez-Melis, Tommi S. Jaakkola, A causal framework for explaining the predictions of black-box sequence-to-sequence models, Proceedings of the conference on Empirical Methods in Natural Language Processing, September 2017.

- Gary van Vuuren, Riaan de Jongh, Tanja Verster, The Impact Of PD-LGD Correlation On Expected Loss And Economic Capital, Economics Research Journal, June 2017.

- pixabay.com

- Clive W.J. Granger, Mark Machina, Forecasting and Decision Theory, in Handbook of Economic Forecasting 1 · December 2006.

- Huan Xu, Constantine Caramanis, Shie Mannor, Sparse Algorithms are not Stable: A No-free-lunch Theorem, IEEE Transactions on Pattern Analysis and Machine Intelligence, January 2012.

APPENDIX

WOE logistic regression

- Data Exploration with Weight of Evidence and Information Value in R

$$\log \frac{P(Y=1|X_j)}{P(Y=0|X_j)} = \underbrace{\log \frac{P(Y=1)}{P(Y=0)}}_{\text{sample log-odds}} + \underbrace{\log \frac{f(X_j|Y=1)}{f(X_j|Y=0)}}_{\text{WOE}},$$

$$\log \frac{P(Y=1|x_1,\ldots,x_p)}{P(Y=0|x_1,\ldots,x_p)} = \log \frac{P(Y=1)}{P(Y=0)} + \sum_{j=1}^{p} \beta_j \log \frac{f(X_j|Y=1)}{f(X_j|Y=0)}$$

$$\text{WOE}_{ij} = \log \frac{P(X_j \in B_i|Y=1)}{P(X_j \in B_i|Y=0)}$$