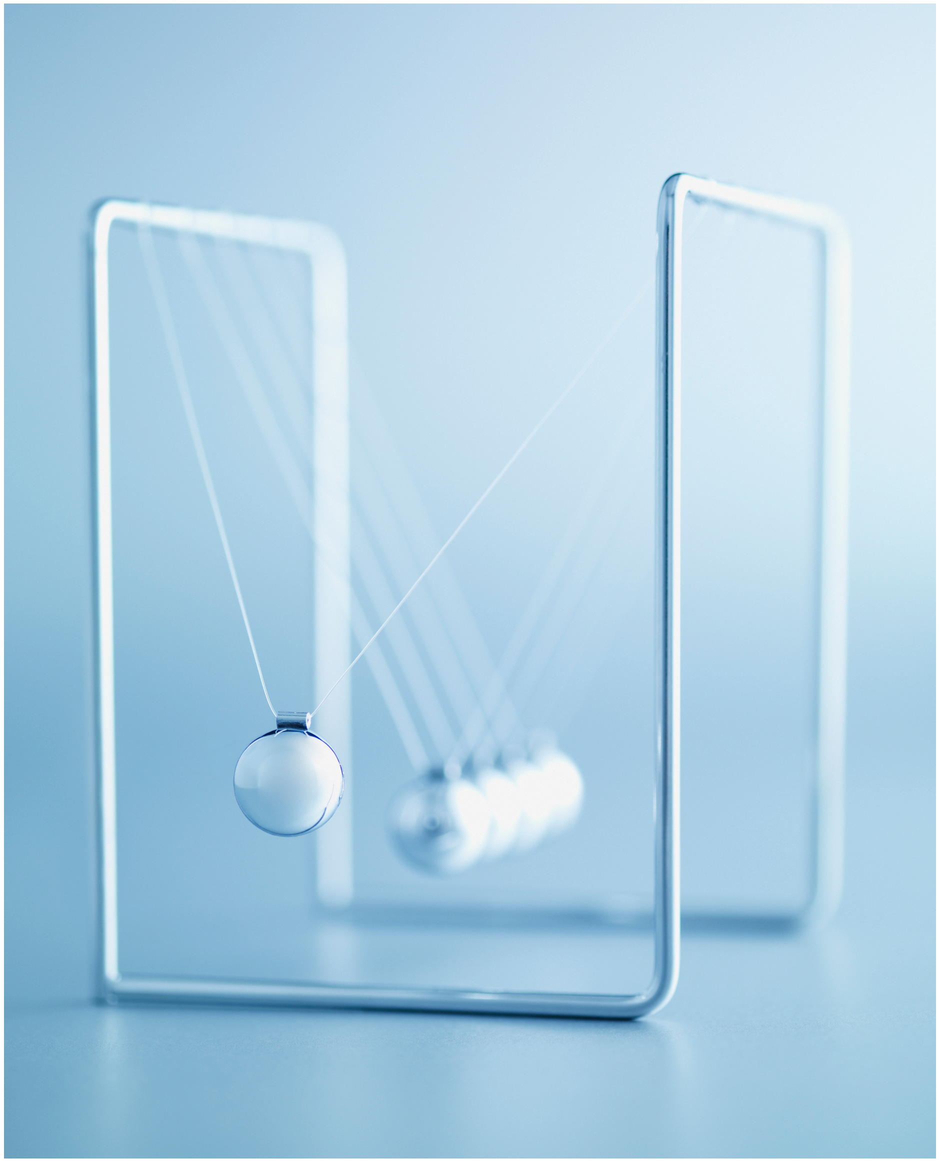


Beyond Correlation

Teasing apart cause and effect to make better business decisions

David Cox
IBM Director, MIT-IBM Watson AI Lab
CTO Forum 2022

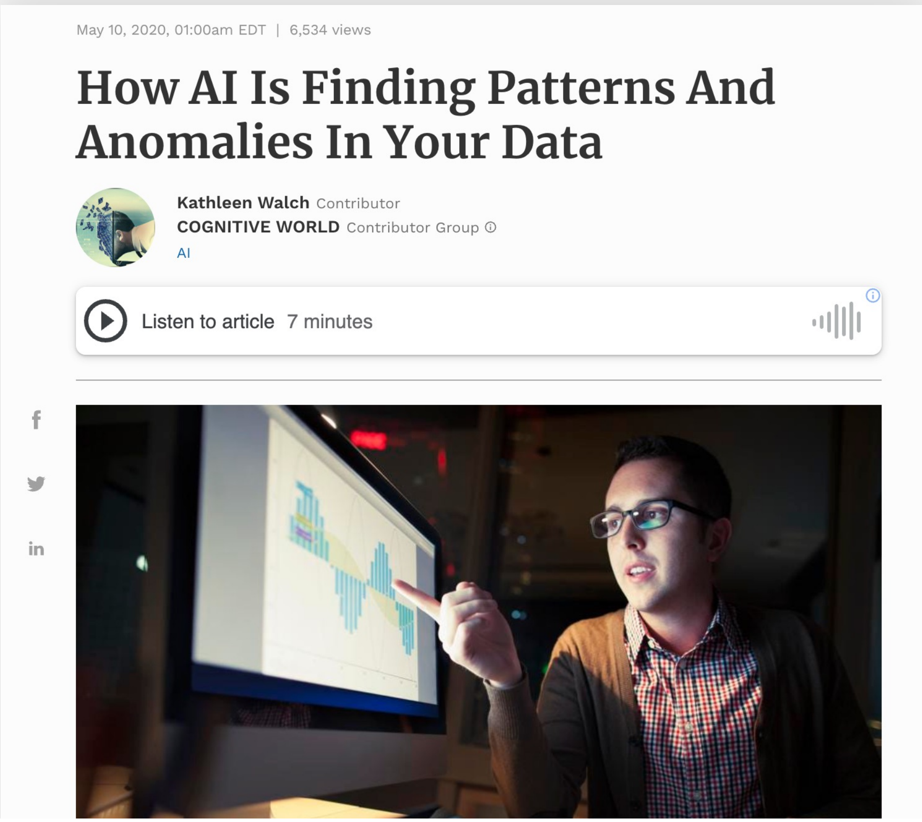
Event
CTO Forum

Date
March 4, 2022

MIT-IBM Watson AI Lab

Confidential information – Use permission granted to CTO Forum.

Most machine learning / deep learning today is fundamentally based on *correlations*


May 10, 2020, 01:00am EDT | 6,534 views

How AI Is Finding Patterns And Anomalies In Your Data

Kathleen Walch Contributor
COGNITIVE WORLD Contributor Group
AI

Listen to article 7 minutes


FEATURE

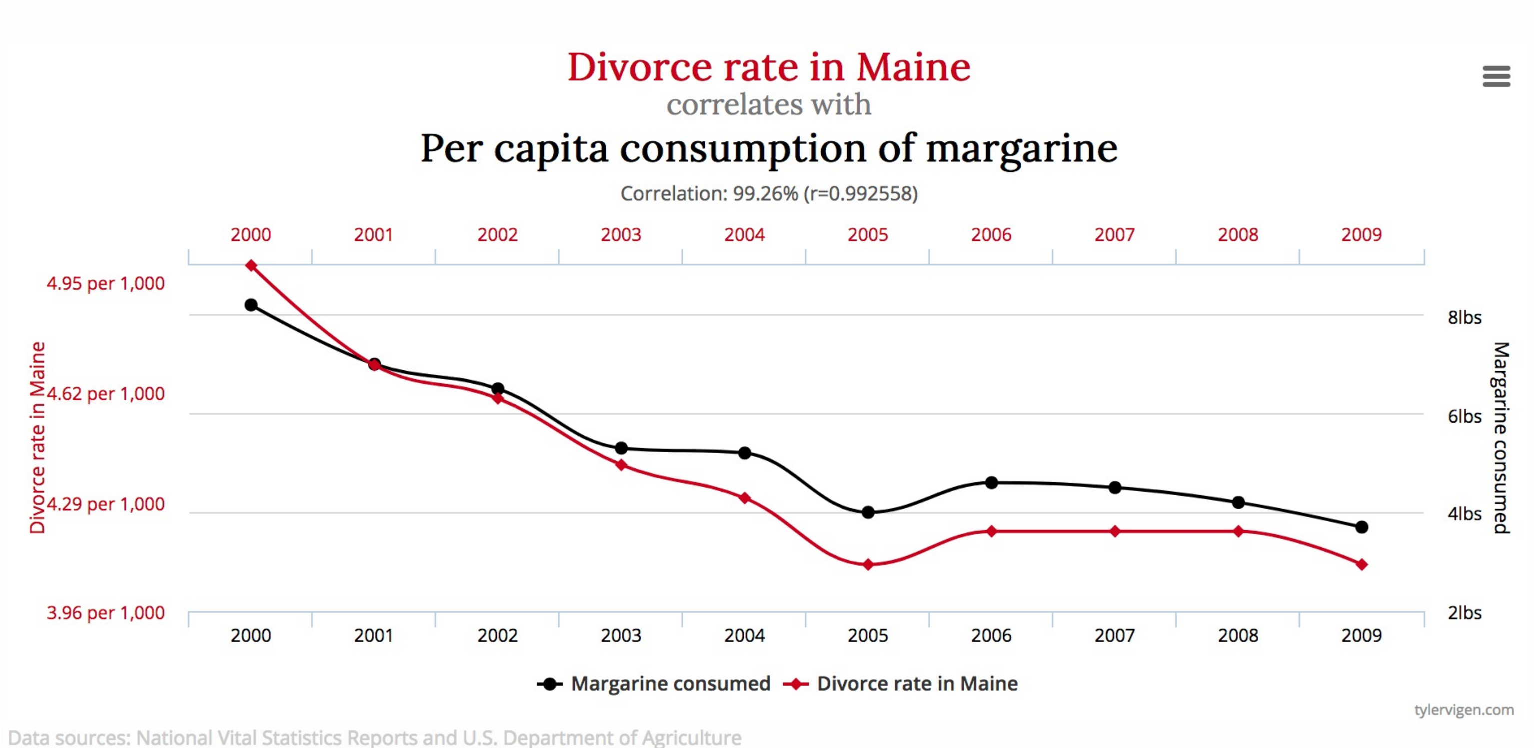
What is data mining? Finding patterns and trends in data

Data mining, sometimes called knowledge discovery, is the process of sifting large volumes of data for correlations, patterns, and trends.

By Thor Olavsrud
Senior Writer, CIO | SEP 27, 2021 2:00 AM PDT

Confidential information – Use permission granted to CTO Forum.

The Problem(s) with Correlation



http://tylervigen.com/spurious-correlations

The Problem(s) with Correlation


thebmj Research Education News & Views Campaigns Archive

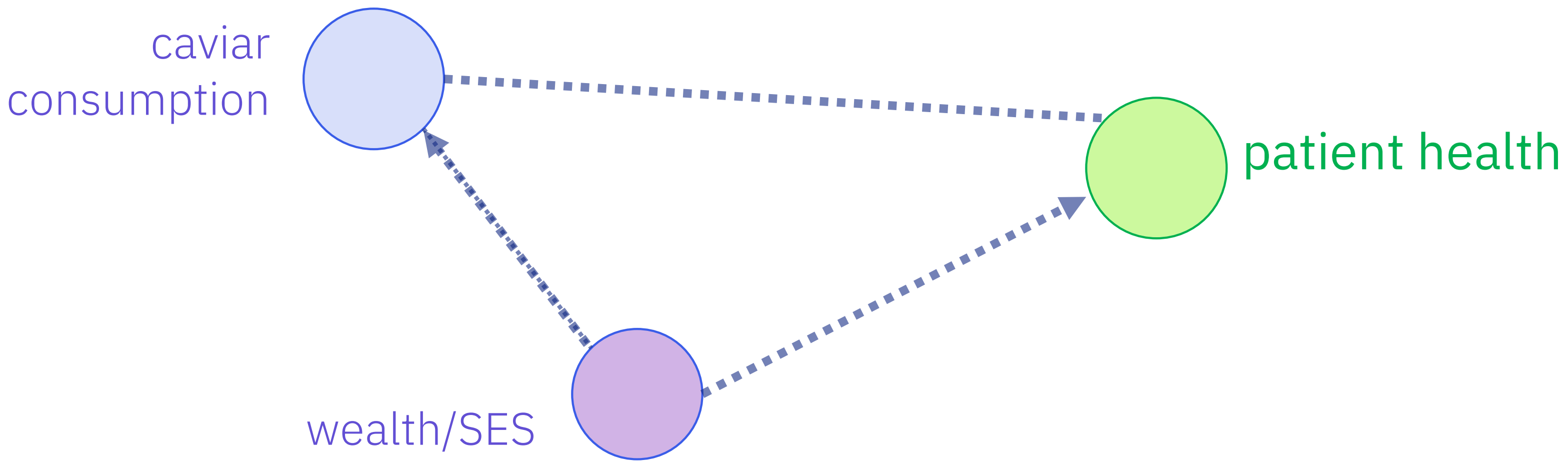
Feature » Christmas 2016: Food for Thought

Is caviar a risk factor for being a millionaire?

BMJ 2016 ; 355 doi: https://doi.org/10.1136/bmj.i6536 (Published 09 December 2016)
Cite this as: *BMJ* 2016;355:i6536

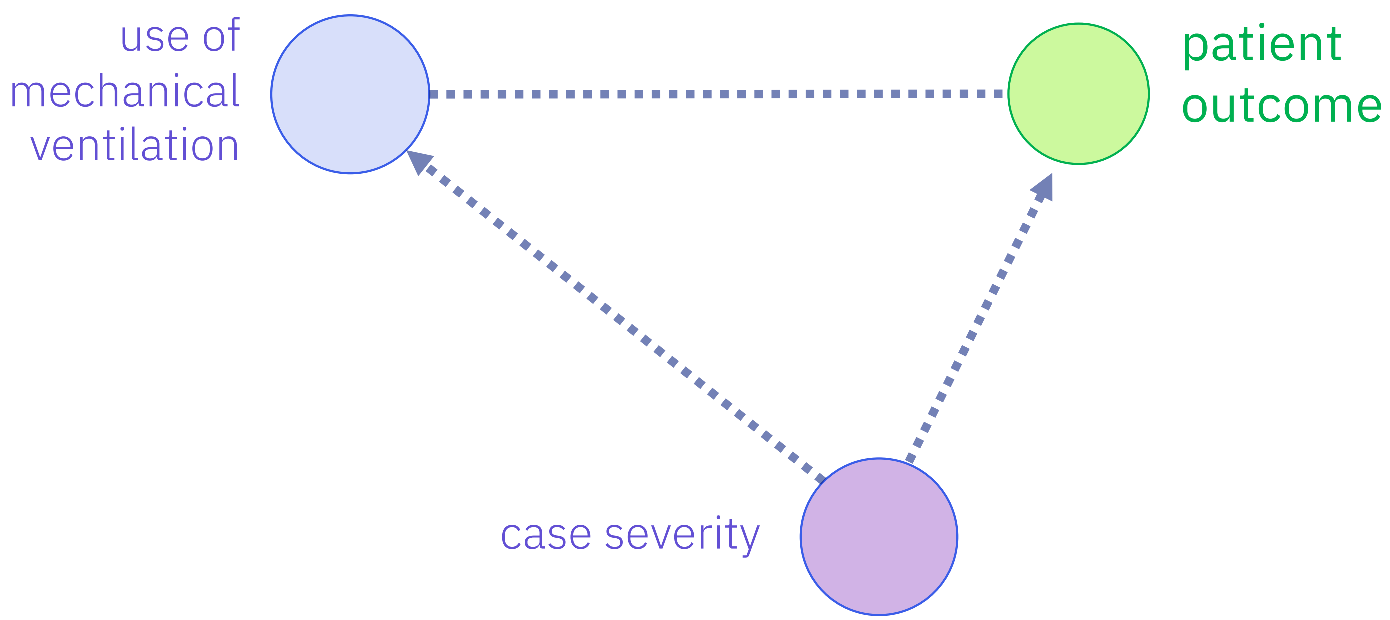
Article | Related content | Metrics | Responses

The tricky business of making decisions



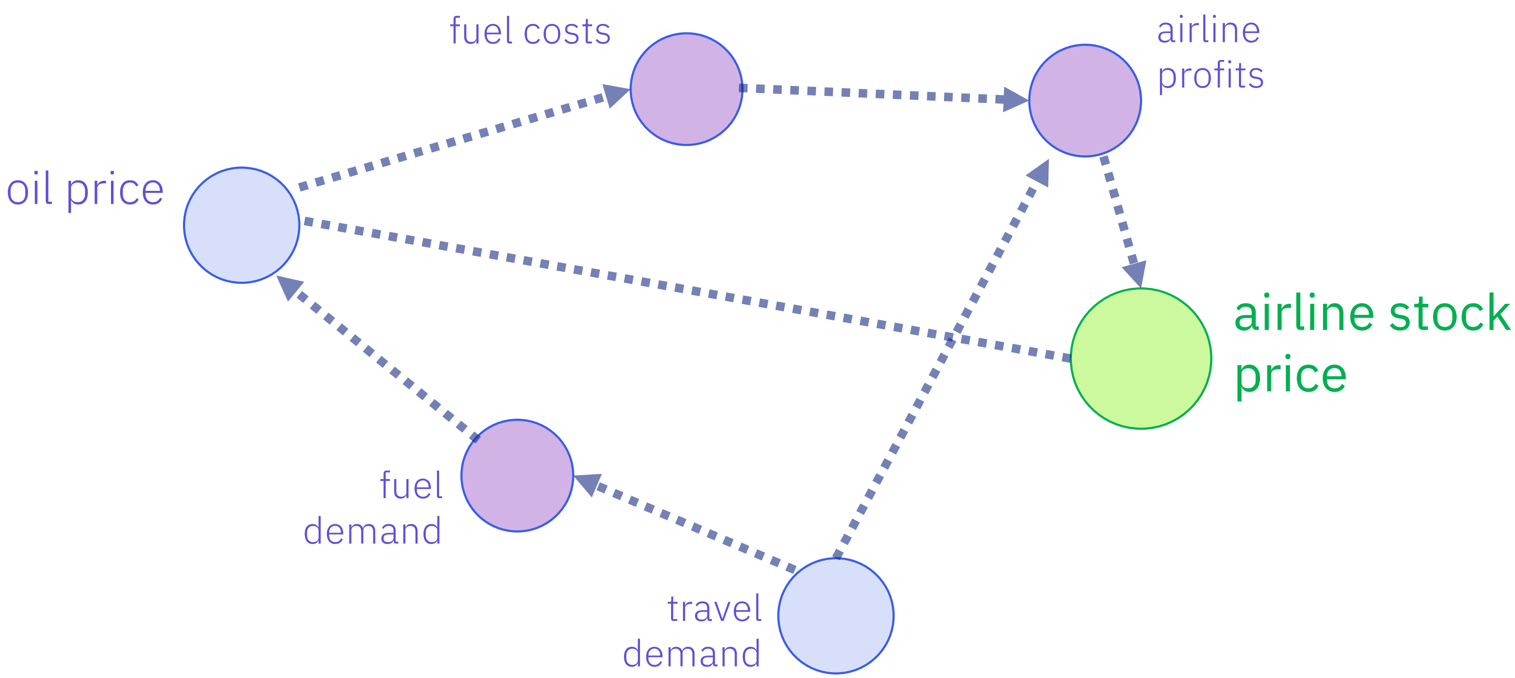
Confidential information – Use permission granted to CTO Forum.

The tricky business of making decisions



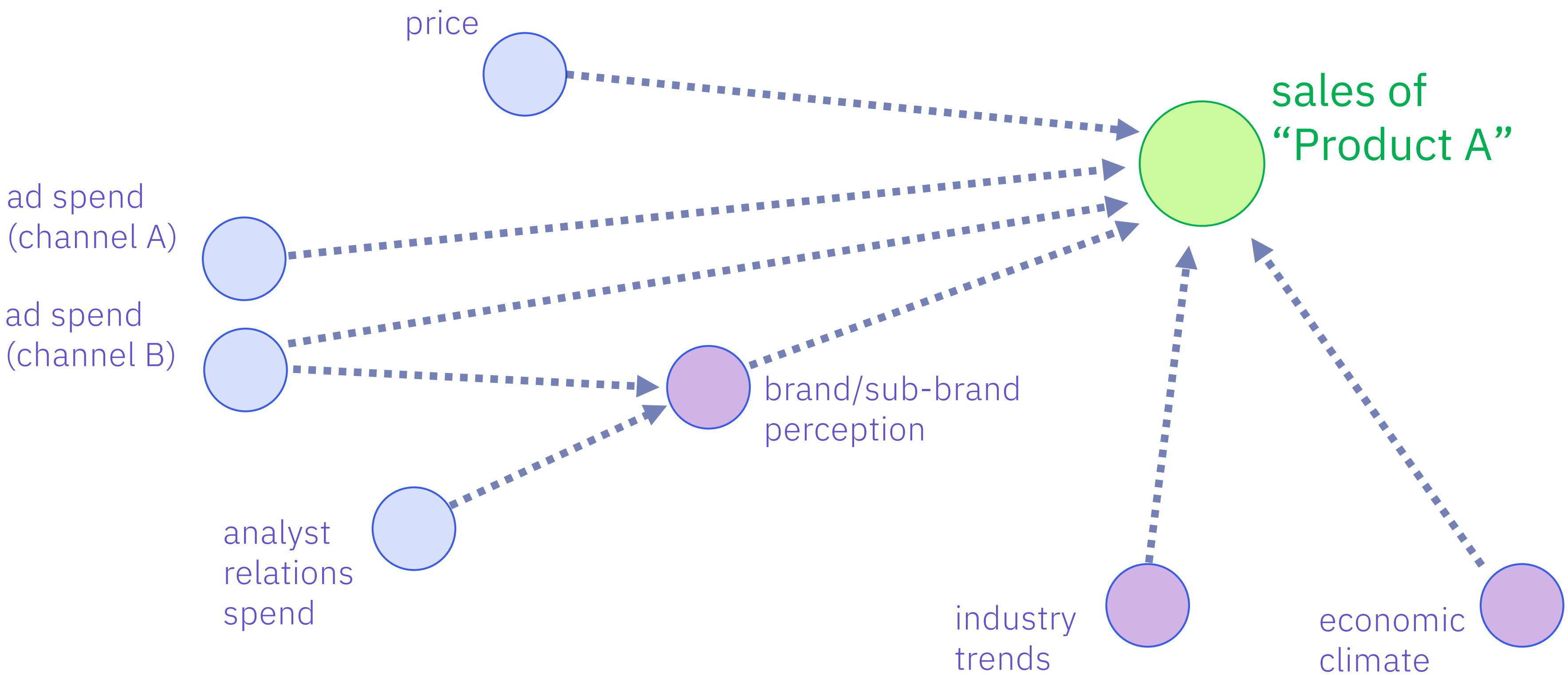
Confidential information – Use permission granted to CTO Forum.

The tricky business of making *business* decisions



Confidential information – Use permission granted to CTO Forum.

The tricky business of making *business* decisions



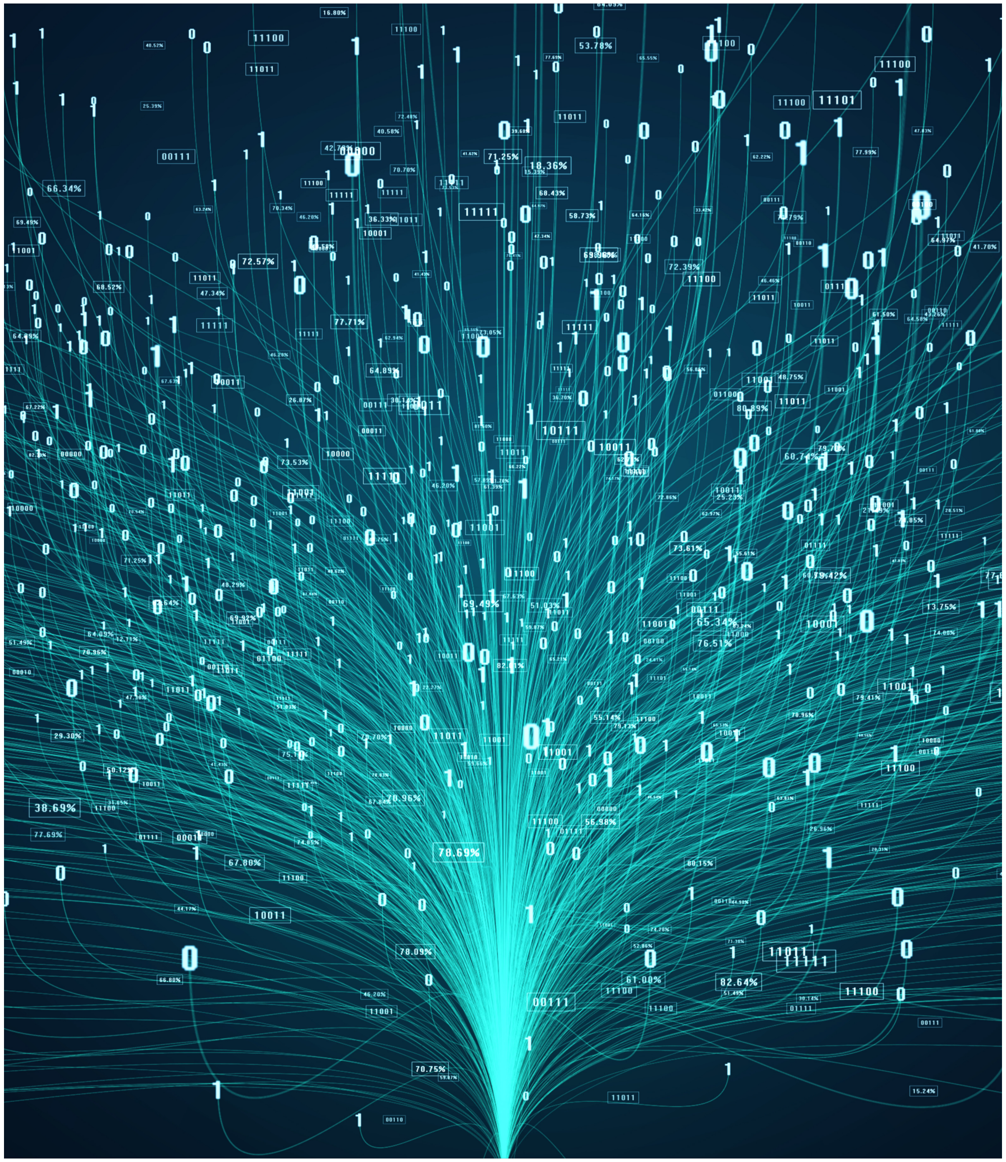
Confidential information – Use permission granted to CTO Forum.

Causal Inference

A subfield of machine learning focused cause and effect relationships, including tools for:

- inferring putative causal structure
- designing experiments/interventions to assess causal structure
- making better decisions when causal structure is known.

Confidential information – Use permission granted to CTO Forum.



A Case Study in Causal Inference for Customer Retention

Research Team

Kristjan Greenewald, Ph.D.
IBM Research

Karthikeyan Shanmugan, Ph.D.
IBM Research

Caroline Uhler, Ph.D.
Institute for Data, Systems, & Society
MIT

MIT-IBM Watson AI Lab

Confidential information – Use permission granted to CTO Forum.

High-Dimensional Featur

Kristjan Greenewald
MIT-IBM Watson AI Lab

Abstract

The estimation of causal treatment
from observational data is a fund
problem in causal inference. To avo
the effect estimator must control for
founders. Hence practitioners often
data for as many covariates as pos
raise the chances of including the
confounders. While this addresses t
this has the side effect of significantly
ing the number of data samples req
accurately estimate the effect due to
creased dimensionality. In this work,
sider the setting where out of a larg
ber of covariates X that satisfy stron
ability, an unknown sparse subset

3. Show that $\hat{\theta}$ is the unique global minimum of the full objective (2).

Step 0: First, we verify a restricted strong convexity condition. Adapted from the $q = 1$ case in [Loh and Wainwright, 2017], we require the following property of the loss function:

Definition 4 (Joint Restricted Strong Convexity (Joint RSC)). *We say a loss $\mathcal{L}_n(\theta)$, $\theta \in \mathbb{R}^{p\times q}$ satisfies an (α, τ) joint RSC condition if for all $\Delta \in \mathbb{R}^{p\times q}$*

$$\langle \nabla \mathcal{L}_n(\theta + \Delta) - \nabla \mathcal{L}_n(\theta), \Delta\rangle \tag{6}$$
$$\geq \begin{cases} \alpha_1 \|\Delta\|_F^2 - \tau_1 \frac{\log p}{n} \|\Delta\|_{1,2}^2 & \|\Delta\|_F \leq 1 \\ \alpha_2 \|\Delta\|_F - \tau_2 \sqrt{\frac{\log p}{n}} \|\Delta\|_{1,2} & \|\Delta\|_F \geq 1. \end{cases}$$

The following is proven in supplement Section 11.

Lemma 3 (Joint RSC for least squares loss). *Assume that $n \geq O(k \log p)$ and $n \geq 4R^2 q \log p$. With high probability (at least $1 - qc_1 \exp(-cn)$), $\mathcal{L}_n$ is (α, τ)-joint RSC for $\alpha_1 = \alpha_2 = \frac{1}{2} \min_j (\lambda_{\min}(\Sigma_x^{(j)}))$ and $\tau_1 = q$, $\tau_2 = \sqrt{q}$. Furthermore, the objective (5) is strongly convex on $\mathbb{R}^S$.*

We also have that with high probability

$$\|\nabla \mathcal{L}_n(\theta^*)\|_{\infty,2} \leq c' \sqrt{\frac{q \log p}{n}}, \tag{7}$$

by applying a norm inequality (2-norm is $\leq \sqrt{q}$ times infinity norm) to the union bounded bound in the proof of Corollary 1 in [Loh and Wainwright, 2015] (the $q = 1$ case) and using $q < p$.

yielding (since $\hat{\Gamma}_{SS}^{(j)}$ is invertible since $n \geq k$ by assumption)

$$\hat{\theta}_{Sj}^{\mathcal{O}} - \theta_{Sj}^* = (\hat{\Gamma}_{SS}^{(j)})^{-1}(-(\hat{\Gamma}_{SS}^{(j)} \theta_{Sj}^* - \hat{\gamma}_S^{(j)}). \tag{10}$$

Appendix D.1.1 of [Loh and Wainwright, 2017] showed that

$$\left\|(\hat{\Gamma}_{SS}^{(j)})^{-1}(\hat{\Gamma}_{SS}^{(j)} \theta_{Sj}^* - \hat{\gamma}_S^{(j)})\right\|_\infty \leq \lambda_{\max}^{1/2}(\Sigma_x^{(j)})\sigma_\epsilon \sqrt{\frac{2\log p}{n}}, \tag{11}$$

with probability at least $1 - c_1'' \exp(-c_2'' \min(k, \log p))$.

Hence we obtain via the union bound that

$$\|\hat{\theta}^{\mathcal{O}} - \theta^*\|_{\infty,\infty} \leq c_3 \sqrt{\frac{\log p}{n}}, \|\hat{\theta}^{\mathcal{O}} - \theta^*\|_{\infty,2} \leq c_3 \sqrt{\frac{q \log p}{n}} \tag{12}$$

with probability at least $1 - c_1 \exp(-c_2 \min(k, \log p))$ (since $k > \log q$ and $p > q$) where c_1, c_2, c_3 are constants.

Now we have the following result, proved in supplement Section 13.

Lemma 5. *Suppose ρ_λ is (μ, γ) amenable and*

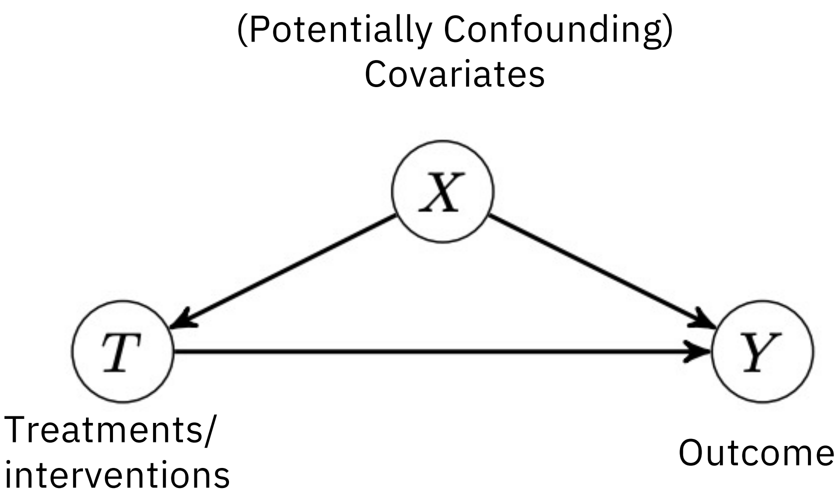
$$\theta_{\min}^* = \min_{i\in S} \|\theta_{i:}^*\|_2 \geq \lambda\gamma + c_3 \sqrt{\frac{\log p}{n}}.$$

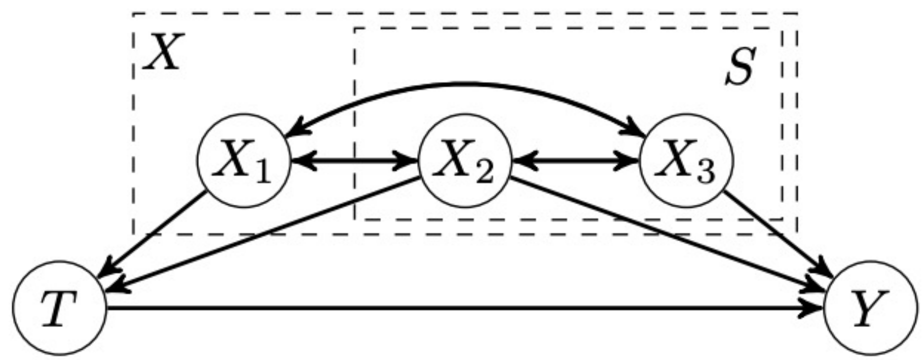
Then with probability at least $1 - c_1 \exp(-c_2 \min(k, \log p))$

$$\lambda \hat{z}_{i:} - \nabla q_\lambda(\|\hat{\theta}_{i:}\|_2) = 0 \quad \forall i \in S.$$

Lemma 5 implies that if $\theta_{\min}^*$ satisfies the given condition, then $\nabla_{\theta_S} \rho_\lambda(\hat{\theta}_S) = 0$, implying that $\hat{\theta}^{\mathcal{O}}$ is a zero

High-dimensional treatment effect estimation



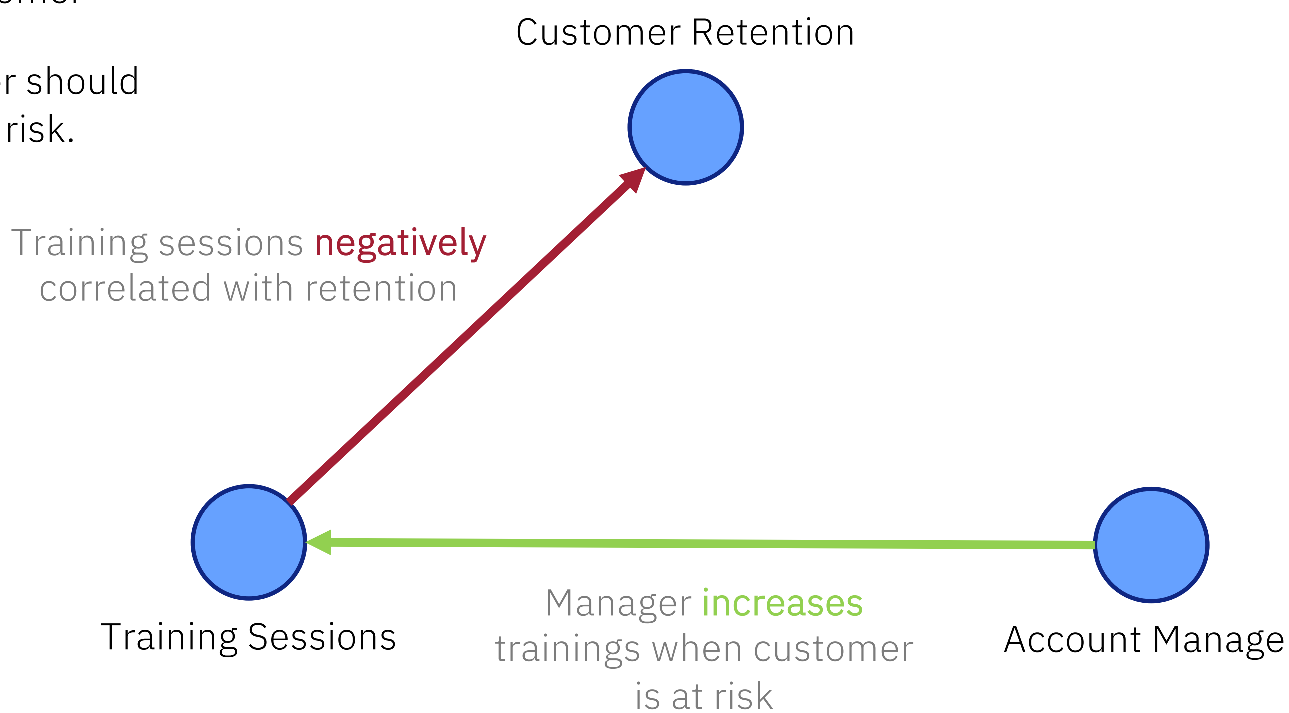


Confidential information – Use permission granted to CTO Forum.

What Factors Drive Customer Retention?

Client Use Case

Discover factors that lead to customer cancellation and identify the top interventions an account manager should take to reduce customer attrition risk.

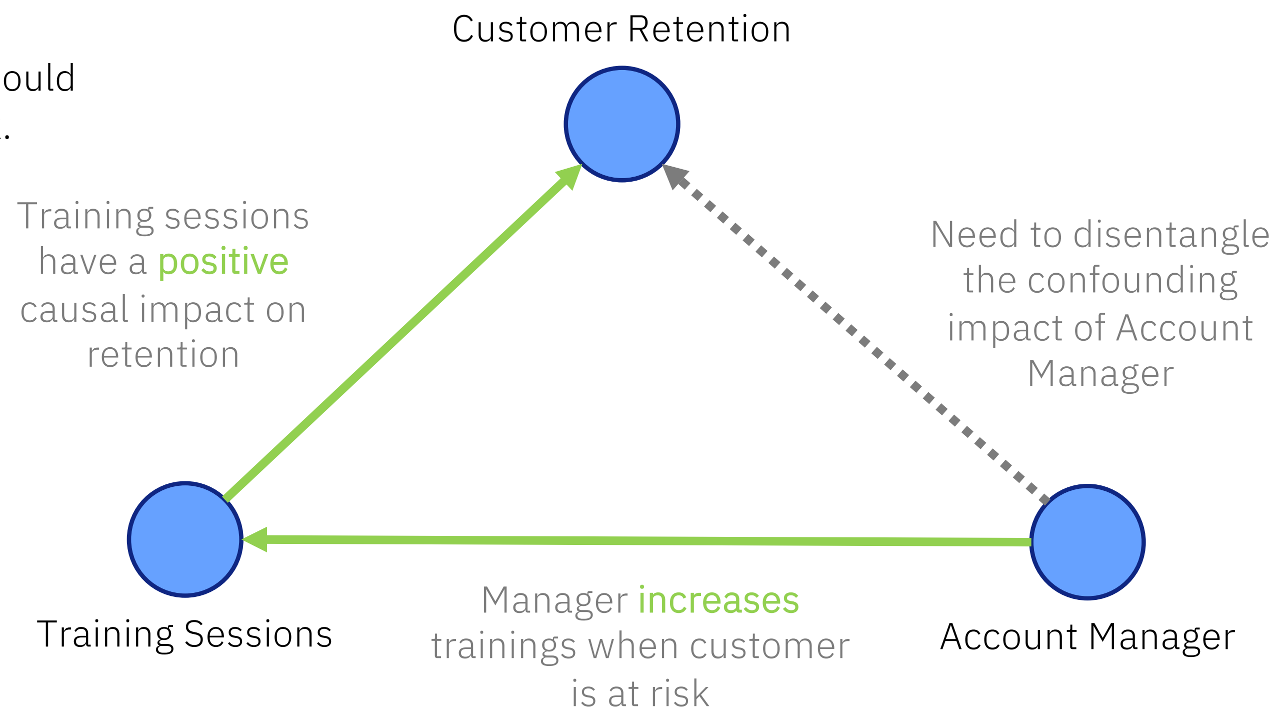


Confidential information – Use permission granted to CTO Forum.

What Factors Drive Customer Retention?

Client Use Case

Discover factors that lead to customer cancellation and identify the top interventions an account manager should take to reduce customer attrition risk.



Confidential information – Use permission granted to CTO Forum.

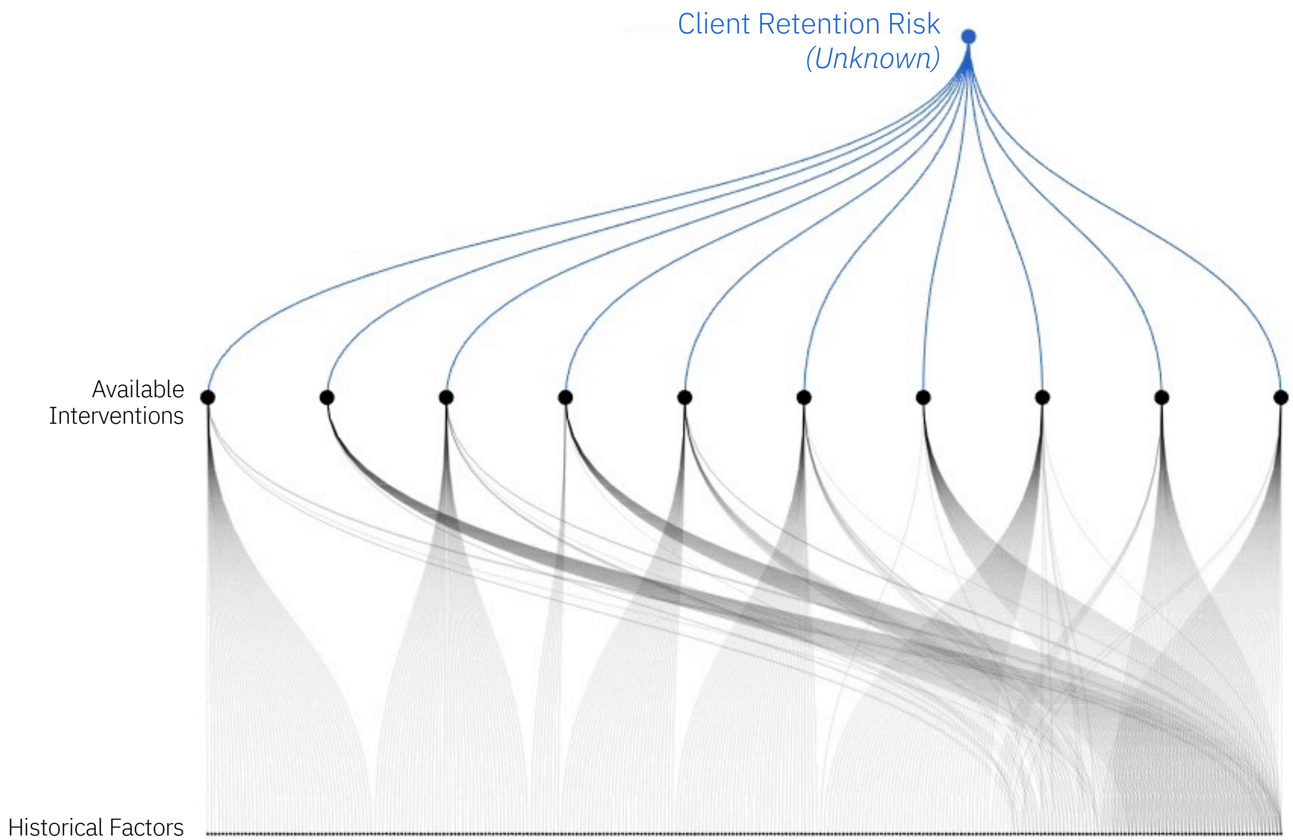
Identifying Causal Structure

Step 1

Discovered the causal structure for a subset of Refinitiv's client risk data.

Insights

Different contextual factors will influence the effectiveness of any given intervention

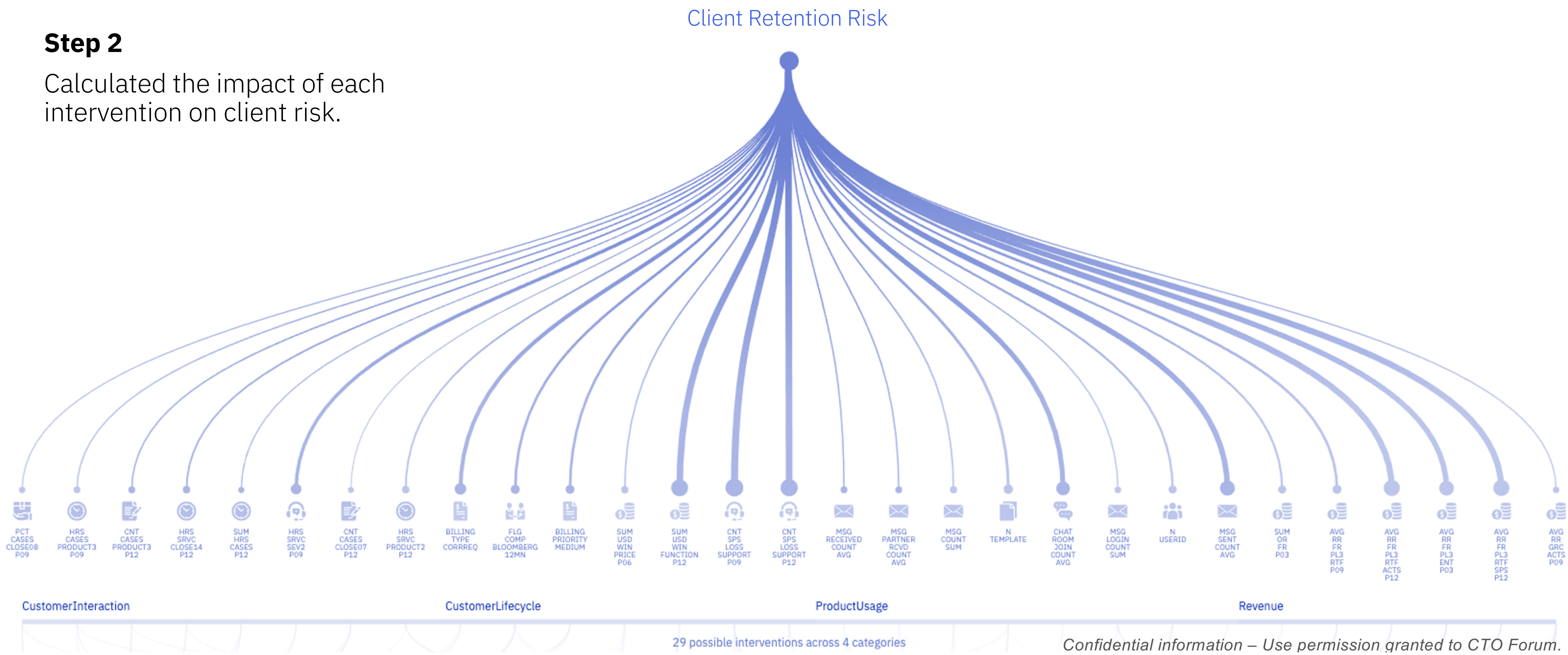


Confidential information – Use permission granted to CTO Forum.

Calculating Intervention Impact

Step 2

Calculated the impact of each intervention on client risk.



Confidential information – Use permission granted to CTO Forum.

Calculating Intervention Impact

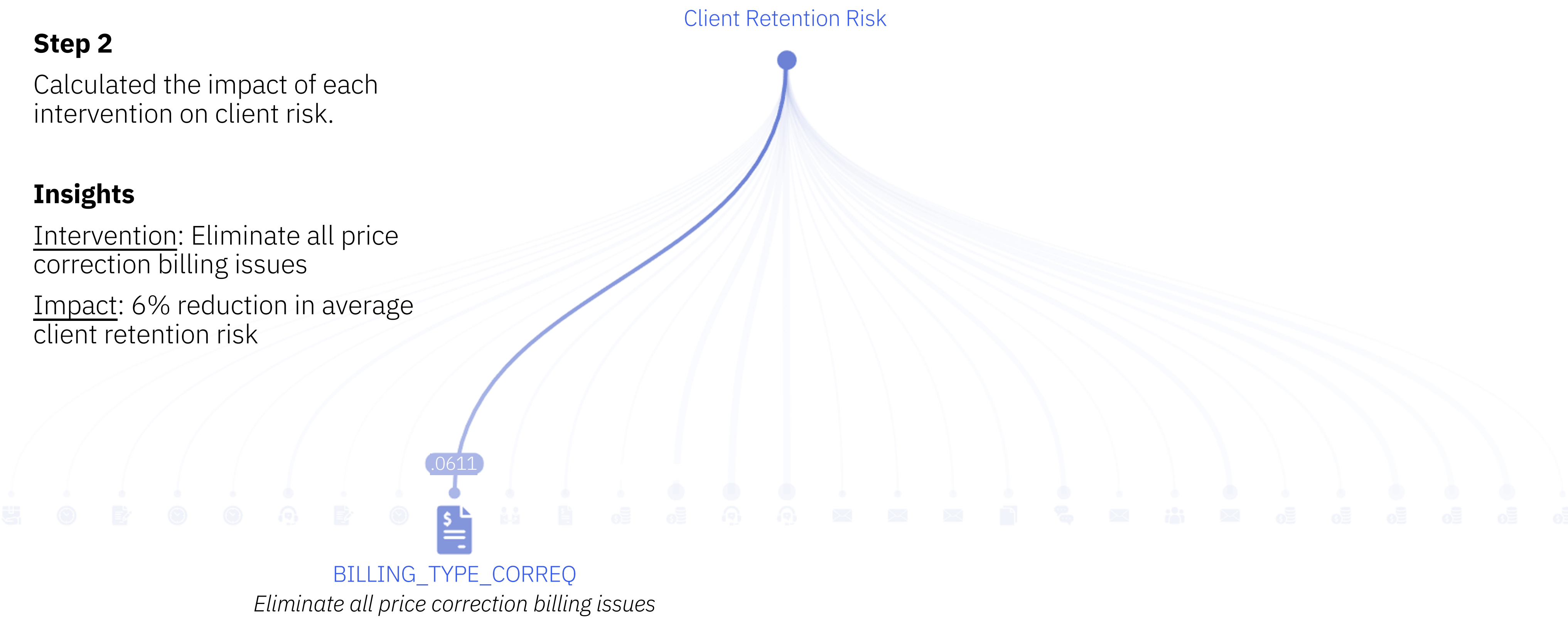
Step 2

Calculated the impact of each intervention on client risk.

Insights

Intervention: Eliminate all price correction billing issues

Impact: 6% reduction in average client retention risk



Confidential information – Use permission granted to CTO Forum.

Calculating Intervention Impact

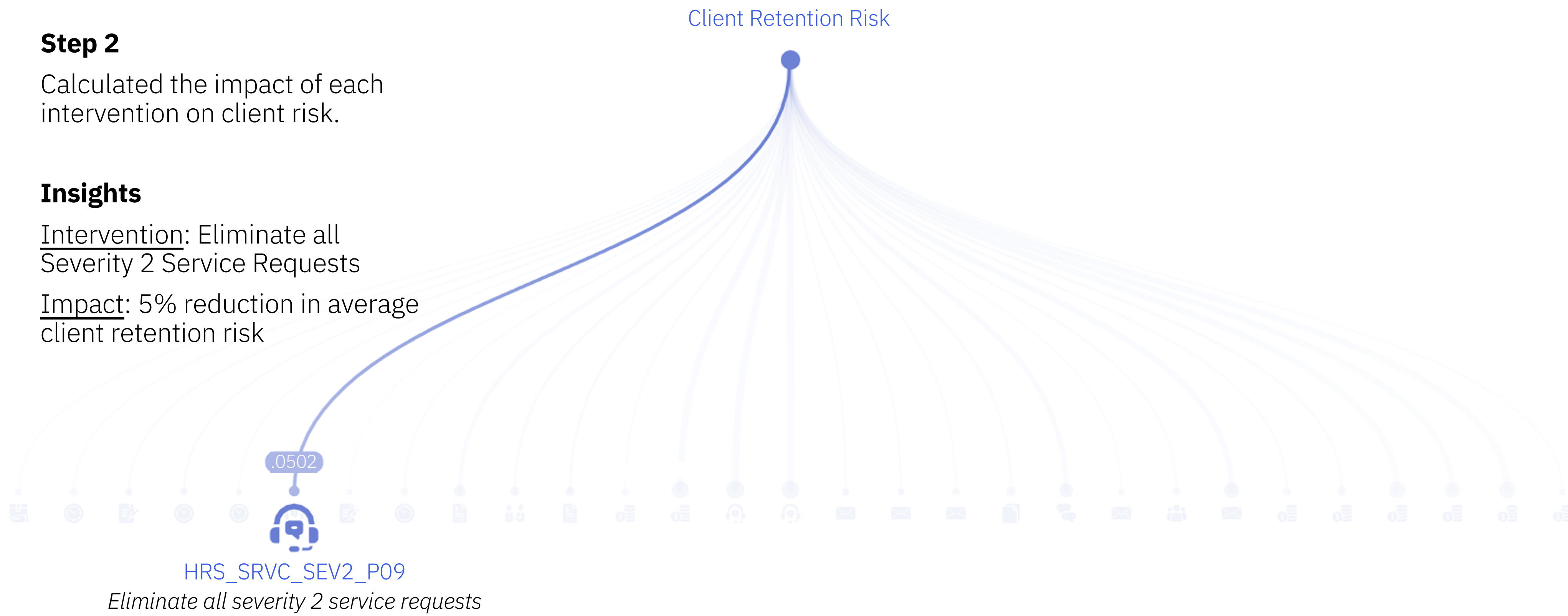
Step 2

Calculated the impact of each intervention on client risk.

Insights

Intervention: Eliminate all Severity 2 Service Requests

Impact: 5% reduction in average client retention risk



Confidential information – Use permission granted to CTO Forum.

Causation vs. Correlation

Step 2

Calculated the impact of each intervention on client risk.

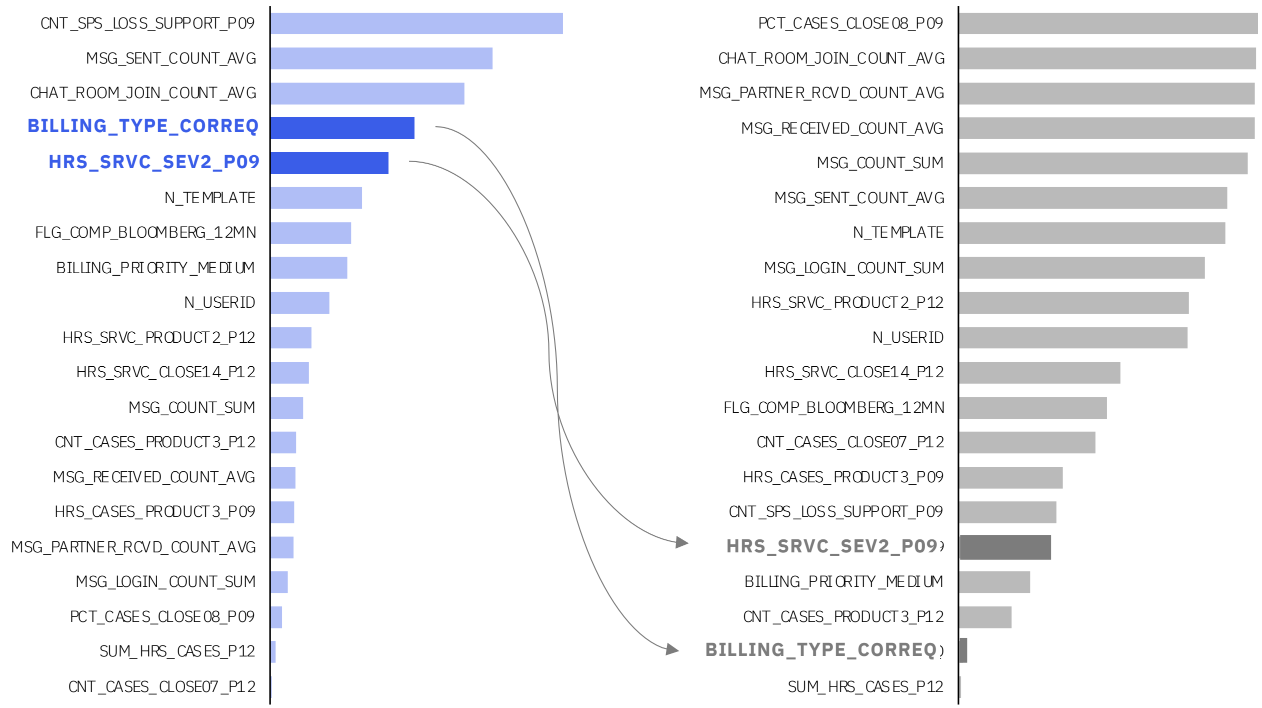
Comparing Correlation to Causation.

Insights

Correlation-based models understate the importance of high impact interventions.

Important Interventions by Causation*

Important Interventions by Correlation*



*Analysis is constrained to a subset of 20 non-revenue related interventional variables

Confidential information – Use permission granted to CTO Forum.

Causation vs. Correlation

Step 2

Calculated the impact of each intervention on client risk.

Compared Correlation to Causation.

Insights

Correlation Model:

If I can increase the number of messages people send via chat, my customer risk will decrease.

Causation Model:

If I can eliminate the need for people to send messages via chat, my customer risk will decrease.

Top 5 Interventions by Causation	Causation	Correlation
CNT_SPS_LOSS_SUPPORT_P09 Eliminate loss of service	0.13	-0.02
MSG_SENT_COUNT_AVG Eliminate need for chat room messages	0.08	-0.02
CHAT_ROOM_JOIN_COUNT_AVG Eliminate need for chat room joins	0.07	-0.05
BILLING_TYPE_CORRREQ Eliminate price correction billing issues	0.04	0.00
HRS_SRVC_SEV2_P09 Eliminate severity 2 service requests	0.04	0.02

Intervention decreases risk

Intervention increases risk

Confidential information – Use permission granted to CTO Forum.

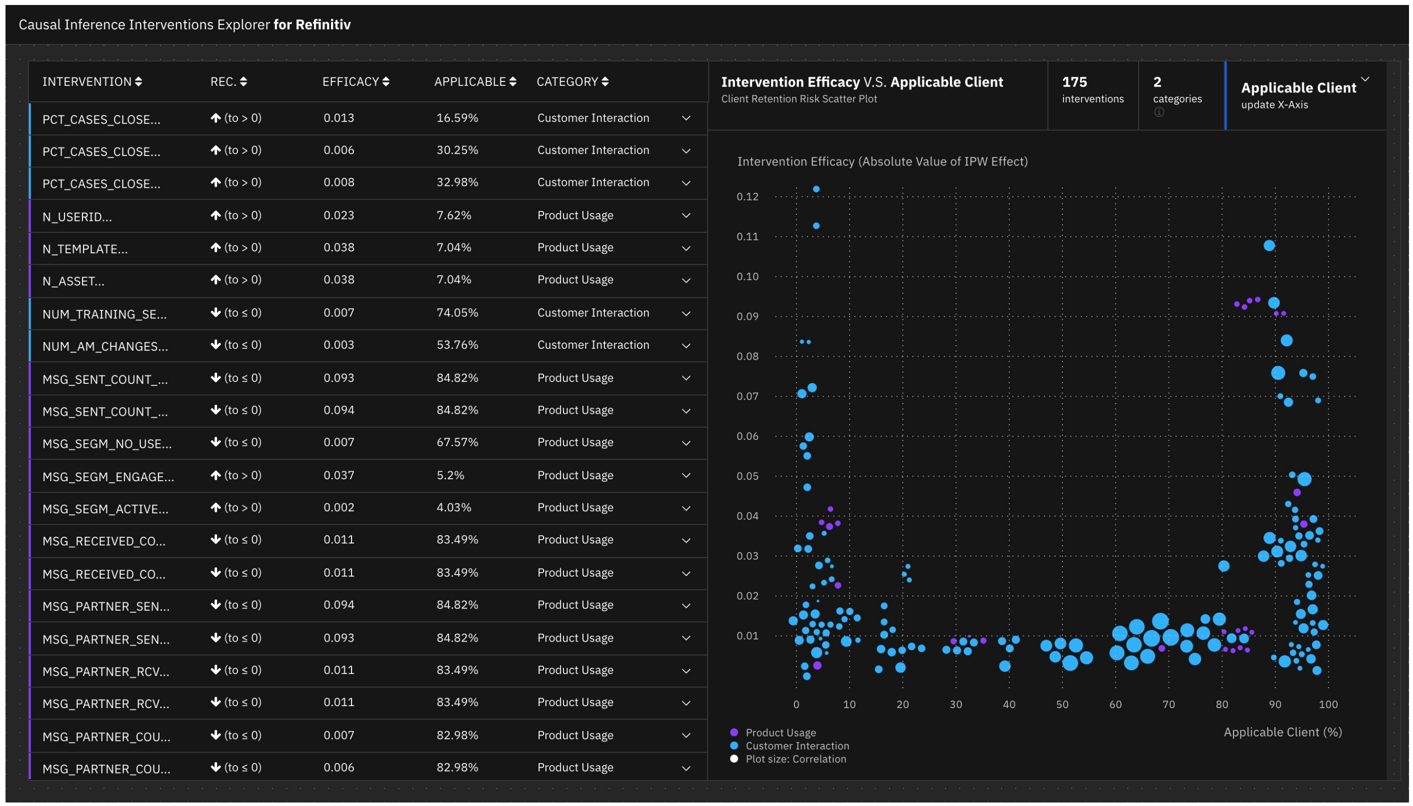
Prioritizing Interventions

Step 3

Applied results to 175 high priority interventions, and created prioritization matrix.

Insights

There can be significant tradeoff between the most impactful interventions, and the percent of the population where the intervention can be applied.

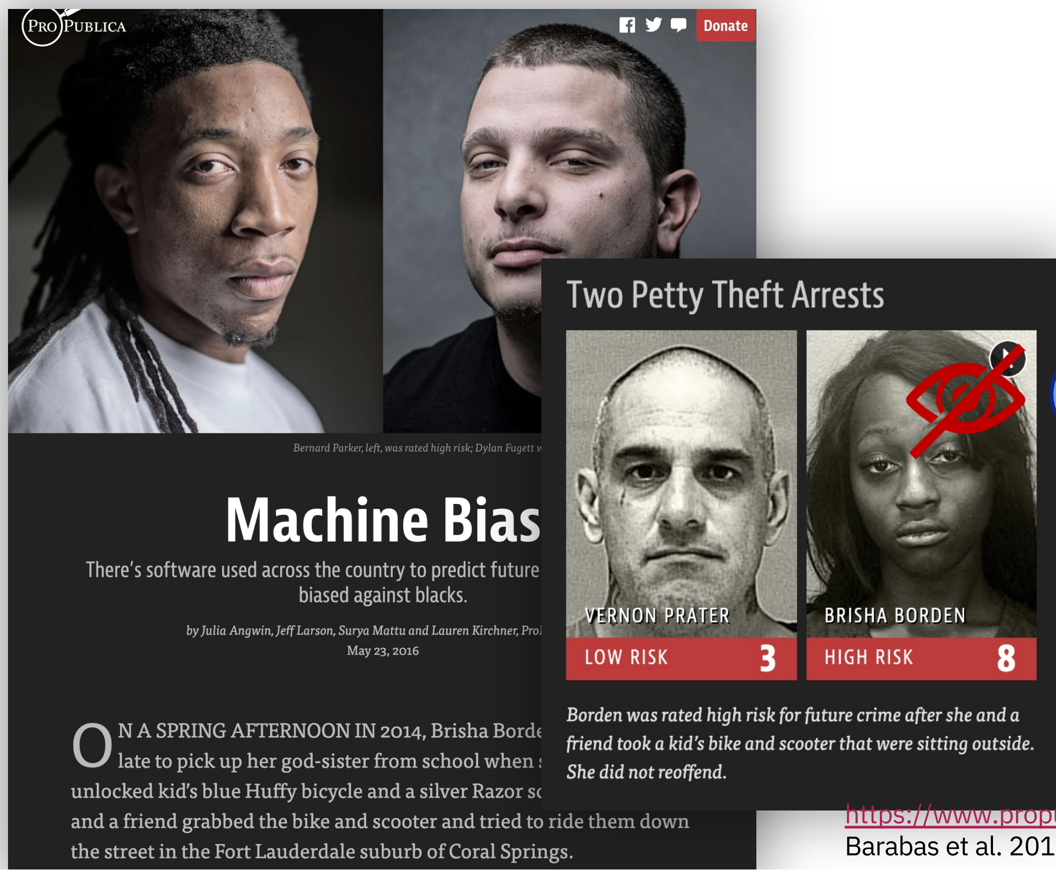


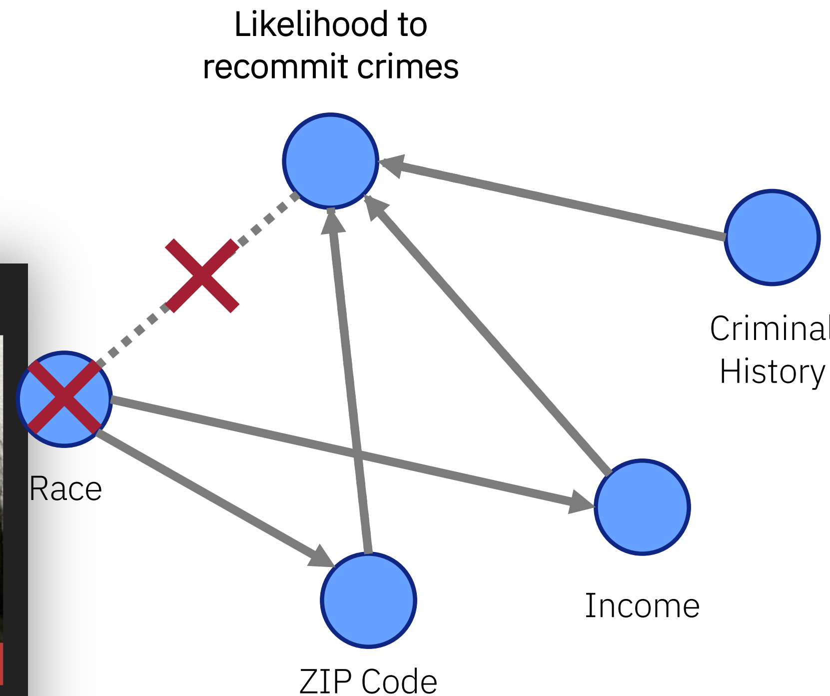
Causal Inference Interventions Explorer **for Refinitiv**

INTERVENTION	REC.	EFFICACY	APPLICABLE	CATEGORY
PCT_CASES_CLOSE...	↑ (to > 0)	0.013	16.59%	Customer Interaction
PCT_CASES_CLOSE...	↑ (to > 0)	0.006	30.25%	Customer Interaction
PCT_CASES_CLOSE...	↑ (to > 0)	0.008	32.98%	Customer Interaction
N_USERID...	↑ (to > 0)	0.023	7.62%	Product Usage
N_TEMPLATE...	↑ (to > 0)	0.038	7.04%	Product Usage
N_ASSET...	↑ (to > 0)	0.038	7.04%	Product Usage
NUM_TRAINING_SE...	↓ (to ≤ 0)	0.007	74.05%	Customer Interaction
NUM_AM_CHANGES...	↓ (to ≤ 0)	0.003	53.76%	Customer Interaction
MSG_SENT_COUNT_...	↓ (to ≤ 0)	0.093	84.82%	Product Usage
MSG_SENT_COUNT_...	↓ (to ≤ 0)	0.094	84.82%	Product Usage
MSG_SEGM_NO_USE...	↓ (to ≤ 0)	0.007	67.57%	Product Usage
MSG_SEGM_ENGAGE...	↑ (to > 0)	0.037	5.2%	Product Usage
MSG_SEGM_ACTIVE...	↑ (to > 0)	0.002	4.03%	Product Usage
MSG_RECEIVED_CO...	↓ (to ≤ 0)	0.011	83.49%	Product Usage
MSG_RECEIVED_CO...	↓ (to ≤ 0)	0.011	83.49%	Product Usage
MSG_PARTNER_SEN...	↓ (to ≤ 0)	0.094	84.82%	Product Usage
MSG_PARTNER_SEN...	↓ (to ≤ 0)	0.093	84.82%	Product Usage
MSG_PARTNER_RCV...	↓ (to ≤ 0)	0.011	83.49%	Product Usage
MSG_PARTNER_RCV...	↓ (to ≤ 0)	0.011	83.49%	Product Usage
MSG_PARTNER_COU...	↓ (to ≤ 0)	0.007	82.98%	Product Usage
MSG_PARTNER_COU...	↓ (to ≤ 0)	0.006	82.98%	Product Usage

Confidential information – Use permission granted to CTO Forum.

Causal Inference for Fairness

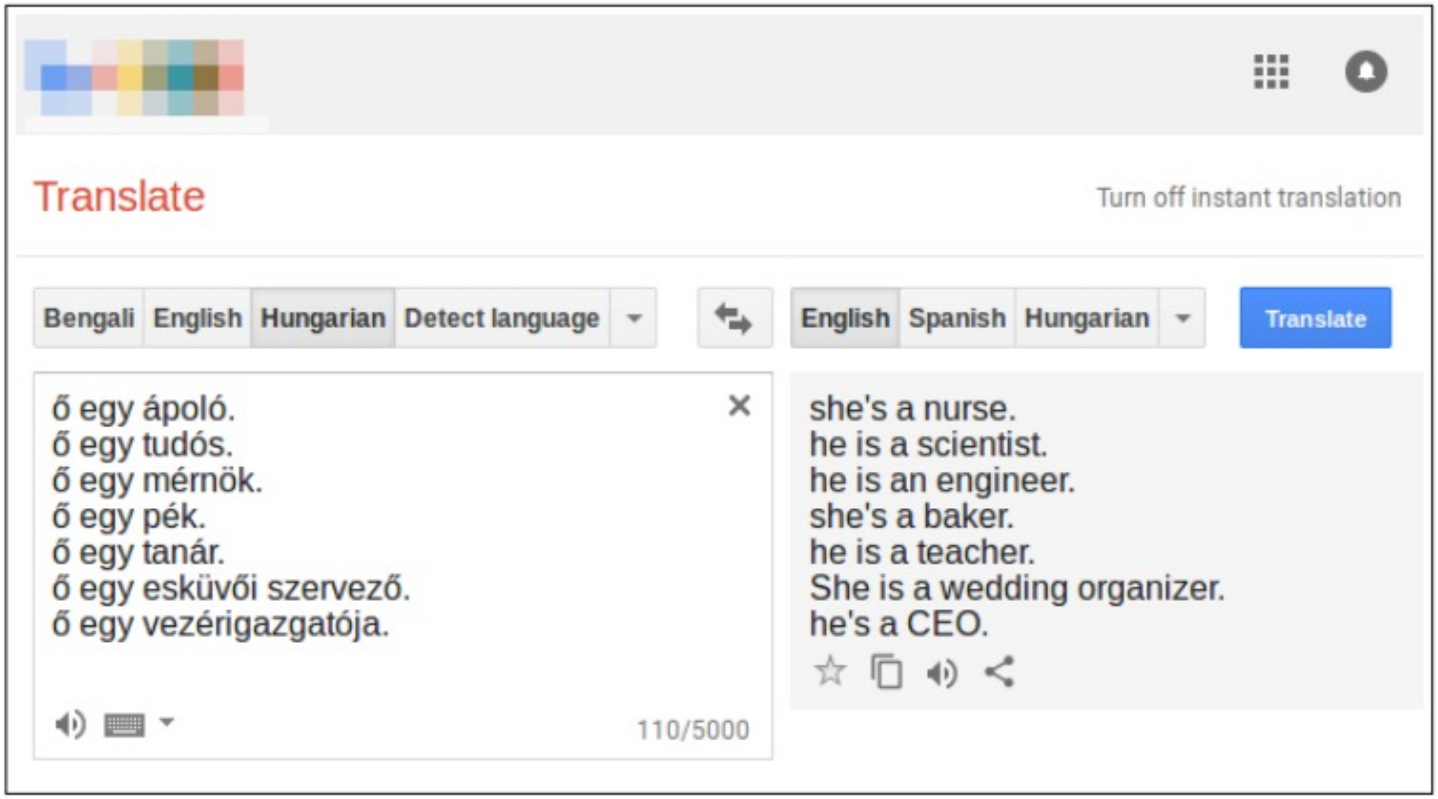




https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
Barabas et al. 2017 PMLR 81:62-76, 2018

Confidential information – Use permission granted to CTO Forum.

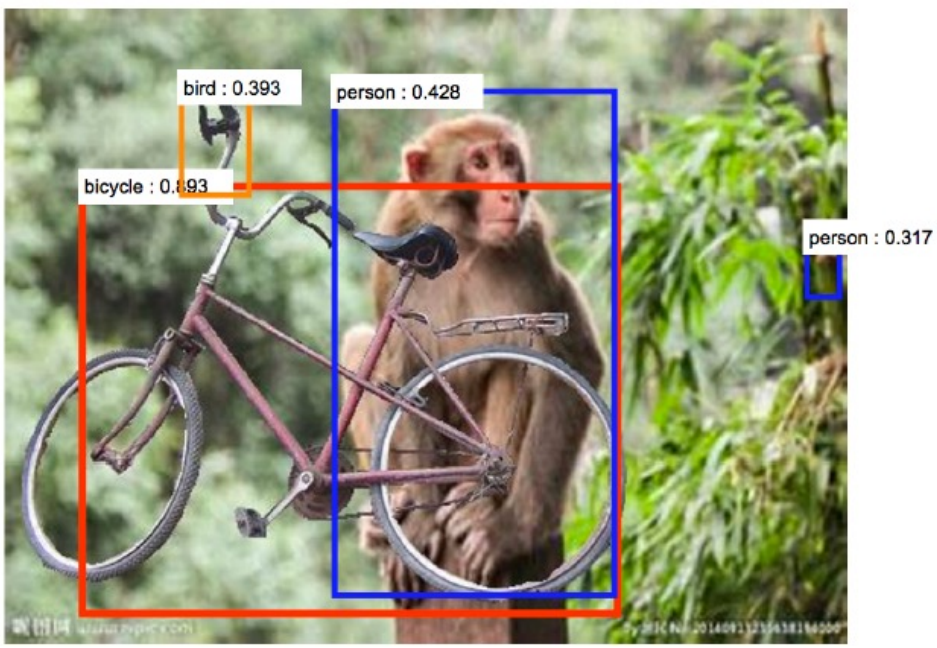
Algorithmic Bias is Widespread

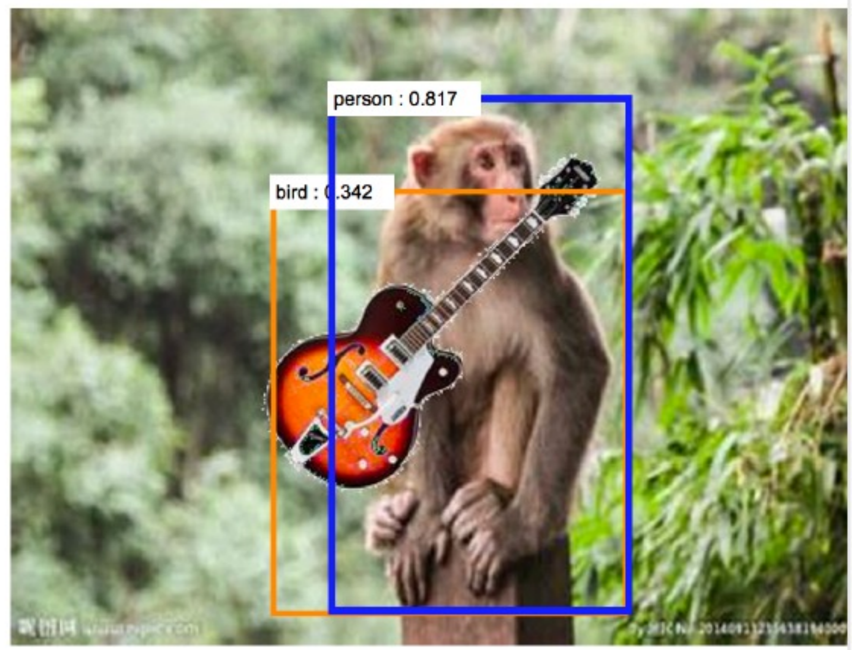


Pratas et al. 2019 "Assessing Gender Bias in Machine Translation – A Case Study with Google Translate"

Confidential information – Use permission granted to CTO Forum.

Algorithmic Bias is Widespread

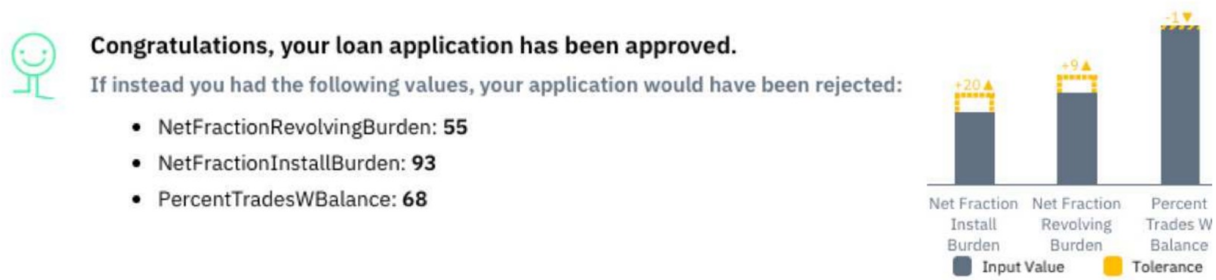




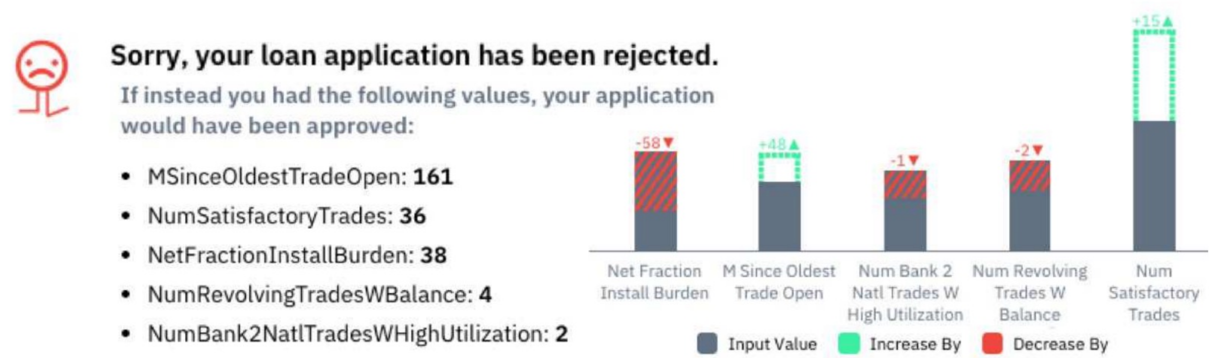
Wang et al. 2017 “Visual concepts and compositional voting”

Confidential information – Use permission granted to CTO Forum.

Causal Explainability



(a) Positive counterfactual explanation



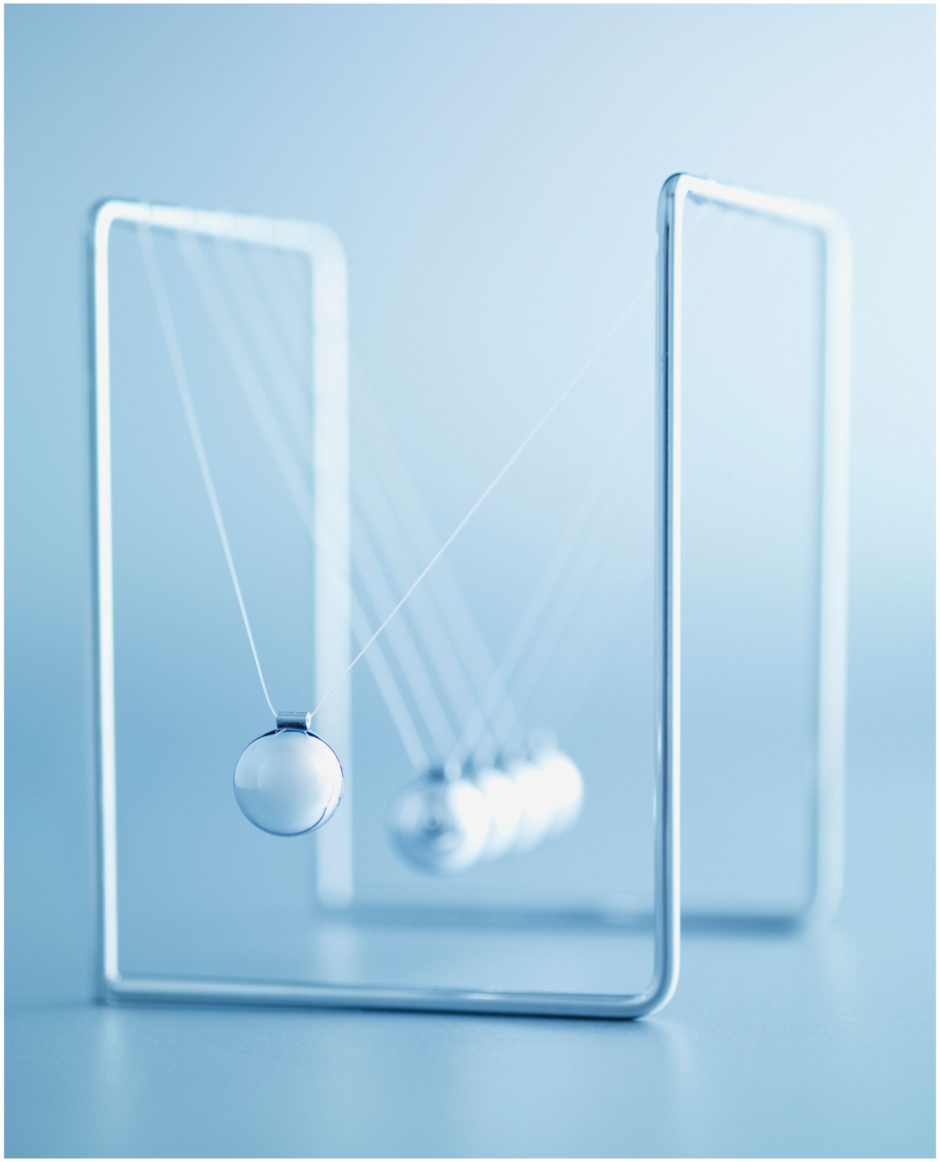
(b) Counterfactual explanation

Counterfactual Explanations:

"the outcome would have been different if the following were true"

McGrath et al. 2018 "Interpretable Credit Application Predictions With Counterfactual Explanations"

Confidential information – Use permission granted to CTO Forum.



Beyond Correlation

Teasing apart cause and effect to make better business decisions

David Cox
IBM Director, MIT-IBM Watson AI Lab
CTO Forum 2022

Event
CTO Forum

Date
March 4, 2022

MIT-IBM Watson AI Lab

Confidential information – Use permission granted to CTO Forum.