Explainable AI
Marketing and retails

Mannheim, 11.01.2019

Agenda

01 What is Explainable AI – XAI ? Why we need it? Overview of methods.

02 XAI for programmatic advertising.

03 XAI for marketing mix modelling.

04 XAI in retail.

05 Looking forward.

XAI in media.

CIO JOURNAL

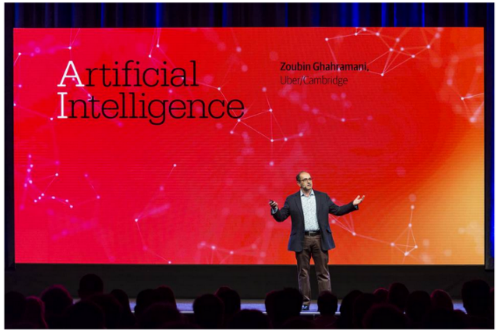
Companies Grapple With AI's Opaque Decision-Making Process

Uber, Xerox's PARC, Capital One among organizations investigating how AI solves problems

By *Sara Castellanos*

May 2, 2018 2:15 pm ET

0 COMMENTS



Zoubin Ghahramani, chief scientist at Uber, speaks at an AI conference this week hosted by OReilly Media Inc. and Intel Corp.s AI division. PHOTO: TRICIA O'NEILL, COURTESY OF O'REILLY MEDIA

Recommended Videos

1. Elon Musk Appears to Smoke Marijuana
2. Powerful Political Messages Sent in Kavanaugh Hearings
3. Naked Body Scanner Reviewed by... Naked Cowboy
4. Three ways 5G technology will change your life

CIO JOURNAL

Capital One Pursues 'Explainable AI' to Guard Against Bias in Models

The effort aims to better understand how a machine-learning model comes to a logical conclusion.

By *Sara Castellanos*

Dec 6, 2016 1:33 pm ET

1 COMMENTS

Adam Wenchel, vice president of data innovation at Capital One Financial Corp., at the AI Summit in New York on Dec. 1. PHOTO: SARA CASTELLANOS / WSJ

Most Popular Videos

1. Elon Musk Appears to Smoke Marijuana
2. How 'Fear' and New York Times Op-Ed Could Impact White House
3. Why Warren Buffett Said No to Lehman and AIG in 2008
4. Powerful Political Messages Sent in

XAI in media.

CIO JOURNAL.

Facing Growing Concern Over AI, Tech Firms Call for ‘Responsible’ Development

By *Steven Norton*
Oct 26, 2017 3:14 pm ET

0 COMMENTS

Employees work in front of computers at the Sinovation Ventures headquarters in Beijing, Aug. 15, 2017. PHOTO: GIULIA MARCHI/BLOOMBERG

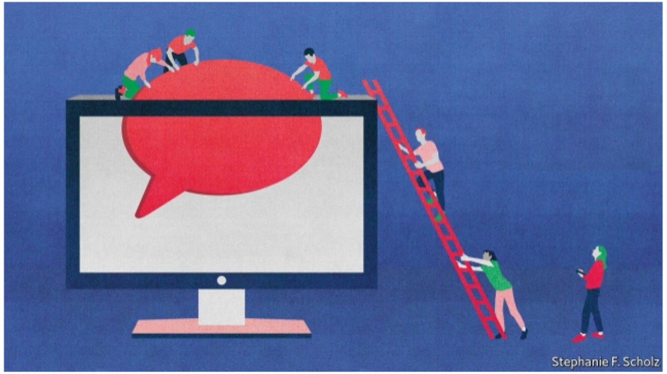
Recommended Videos

1. Elon Musk Appears to Smoke Marijuana
2. Three ways 5G technology will change your life
3. Powerful Political Messages Sent in Kavanaugh Hearings
4. Naked Body Scanner Reviewed by... Naked Cowboy

AI in society

For artificial intelligence to thrive, it must explain itself

If it cannot, who will trust it?



Stephanie F. Scholz

Print edition | Science and technology ›
Feb 15th 2018

AI applications that can make you worry.

The New York Times

Sent to Prison by a Software Program's Secret Algorithms

Chief Justice John G. Roberts Jr., center, recently said that the day of using artificial intelligence in courtrooms was already here, "and it's putting a significant strain on how the judiciary goes about doing things." Stephen Crowley/The New York Times

- Eric L. Loomis, who was sentenced to six years in prison based in part on a private company's proprietary software. Mr. Loomis says his right to due process was violated by a judge's consideration of a report generated by the software's secret algorithm, one Mr. Loomis was unable to inspect or challenge.

AI applications that can make you worry.

TECH | By Jordan Pearson | Feb 2 2017, 4:23pm

AI Could Resurrect a Racist Housing Policy

And why we need transparency to stop it.

SHARE TWEET

Data has always been a weapon. Between 1934 and 1968 the US Federal Housing Administration systematically denied loans to black people by using entire neighbourhoods, colour-coded by perceived risk factor, as their decision-making metric. Modern computer scientists might call this intentionally "coarse" data.

This practice, known as redlining, had damaging financial and social effects that spanned generations of black families. And now, experts worry that similar practices could return in the algorithms that make decisions about who poses a risk to their community, or, rather chillingly, who deserves to be granted a loan.

AI applications that can make you worry.

SCIENCE

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

By Colin Lecher | @colinlecher | Mar 21, 2018, 9:00am EDT

Illustrations by William Joel; Photography by Amelia Holowaty Krales

SHARE

For most of her life, Tammy Dobbs, who has cerebral palsy, relied on her family in Missouri for care. But in 2008, she moved to Arkansas, where she signed up for a state program that provided for a caretaker to give her the help she needed.

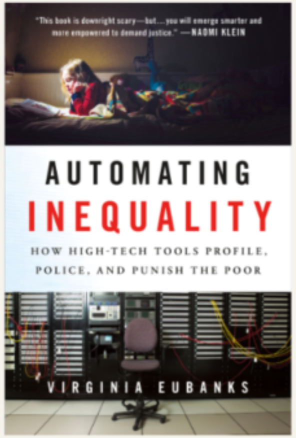
There, under a Medicaid waiver program, assessors interviewed beneficiaries and decided how frequently the caretaker should visit. Dobbs' needs were extensive. Her illness left her in a wheelchair and her hands stiffened. The most basic tasks of life — getting out of bed, going to the bathroom, bathing — required assistance, not to mention the trips to yard sales she treasured. The nurse assessing her situation allotted Dobbs 56 hours of home care visits per week, the maximum allowed under the program.

AI applications that can make you worry.

AUTOMATING INEQUALITY

How High-Tech Tools Profile, Police, and Punish the Poor

Virginia Eubanks
St. Martin's Press



READ AN EXCERPT →

BUY THE BOOK

Hardcover

St. Martin's Press
St. Martin's Press
01/23/2018
ISBN: 9781250074317
272 Pages

$26.99

Amazon | Barnes & Noble | Books-a-Million | IndieBound | Powells

But companies take steps.

Salesforce is hiring its first Chief Ethical and Humane Use officer to make sure its artificial intelligence isn't used for evil

Rosalie Chan Dec. 16, 2018, 3:10 PM

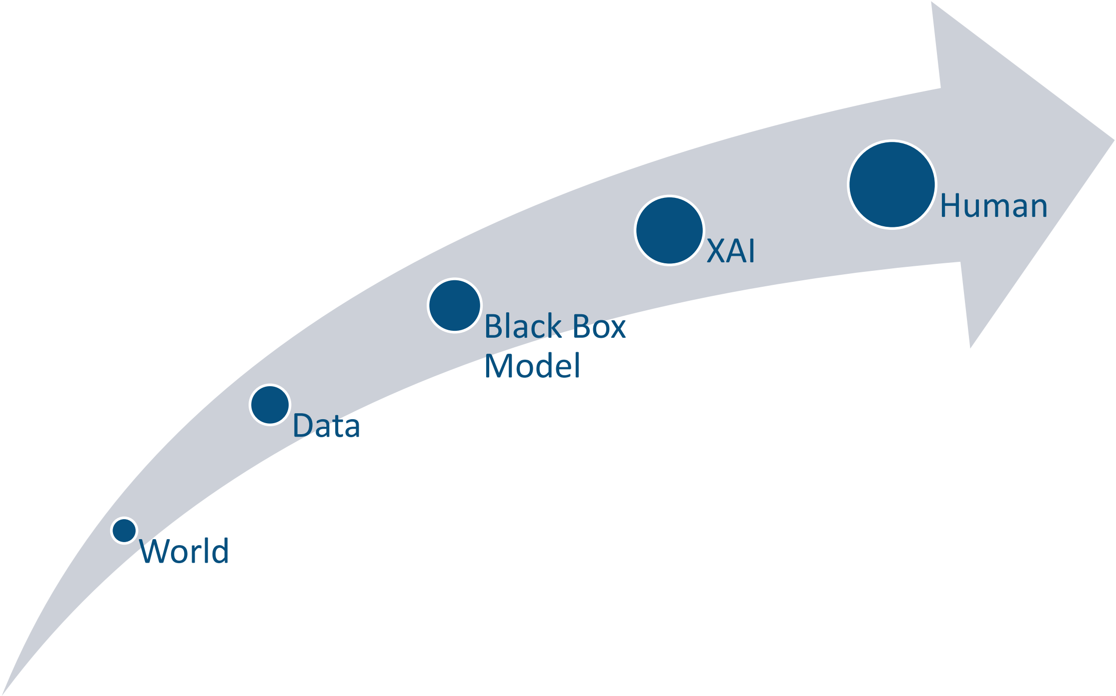
What is XAI ?

Explainable AI

- XAI aims to produce "glass box" models that are explainable to a "human-in-the-loop", without greatly sacrificing AI performance.
- Human users should be able to understand the AI's cognition (both in real-time and after the fact), and should be able to determine when to trust the AI and when the AI should be distrusted

Source https://en.wikipedia.org/wiki/Marketing_mix_modeling

XAI in action.



What Data Scientists get wrong about explainability.

01 Judge AIs as alternatives rather than aides

02 Expect stakeholders to "think more like me"

03 Optimize for model performance over enterprise utility

04 Value XAI only as a placebo

05 Believe what is said is what will be heard

06 Provide a single explanation for all audiences

07 Undervalue explanation friendly features

08 Fail to design for debugging

09 Assume rather than demonstrate generalizability

10 Think moonshots are the model

Source https://xai.world/2018/01/25/what-data-scientists-get-wrong-about-explainability/

Point 1

01 **Judge AIs as alternatives rather than aides**

02 Expect stakeholders to "think more like me"

03 Optimize for model performance over enterprise utility

04 Value XAI only as a placebo

05 Believe what is said is what will be heard

- "Considering machine learning systems as a replacement for a human makes it is easier to assume that no detailed explanation is needed so long as they do their job. However, most AIs will be aides that augment human decision making and clearly need to explain their results."

Point 2

01 Judge AIs as alternatives rather than aides

02 Expect stakeholders to “think more like me”

03 Optimize for model performance over enterprise utility

04 Value XAI only as a placebo

05 Believe what is said is what will be heard

- “Machine learning practitioners must adapt our systems to fit the expectations of the world, rather than expect the world to adapt to our expectations.”

Point 3

01 Judge AIs as alternatives rather than aides

02 Expect stakeholders to “think more like me”

03 **Optimize for model performance over enterprise utility**

04 Value XAI only as a placebo

05 Believe what is said is what will be heard

- “It’s natural for data scientist to see everything as an optimization problem. The trick is knowing what to optimize for. It is tempting to work at the lab bench focused on optimizing your model as measured by the typical machine learning metrics. What is harder to do but far more valuable is to optimize for the benefits and overall utility of the encompassing system.”

Point 4

01 Judge AIs as alternatives rather than aides

02 Expect stakeholders to “think more like me”

03 Optimize for model performance over enterprise utility

04 Value XAI only as a placebo

05 Believe what is said is what will be heard

- “It is easy to imagine the only benefit of explainability is to placate the users. That is a tempting idea, it is also wrong.”

Point 5

01 Judge AIs as alternatives rather than aides

02 Expect stakeholders to "think more like me"

03 Optimize for model performance over enterprise utility

04 Value XAI only as a placebo

05 Believe what is said is what will be heard

- "It is common to have a disconnect between what a machine learning model is actually communicating vs. what the stakeholders are hearing. Users can extrapolate the recognition of a pattern into unwarranted confidence in a presumed course of action."

Point 6

06 Provide a single explanation for all audiences

07 Undervalue explanation friendly features

08 Fail to design for debugging

09 Assume rather than demonstrate generalizability

10 Think moonshots are the model

- “Data scientists tend to think in terms of rigorous ideals. They like clear cut goals and provably correct answers. So it is natural to think that a given model or a specific result has a single definitive explanation. But of course they don’t, different audiences need different explanations.”

Point 7

06 Provide a single explanation for all audiences

07 **Undervalue explanation friendly features**

08 Fail to design for debugging

09 Assume rather than demonstrate generalizability

10 Think moonshots are the model

- “Model features shouldn’t be judged just on statistical equivalence, run time speed and server resource use. They should also be judged based whether they clarify the connection between our model and reality. They should be judged based on how much they contribute to explanations and generalizability.”

Point 8

06 Provide a single explanation for all audiences

07 Undervalue explanation friendly features

08 Fail to design for debugging

09 Assume rather than demonstrate generalizability

10 Think moonshots are the model

- "Every system will have failures in production. Robust systems design in from the start the mechanisms needed to quickly isolate and resolve those failures. For example, if we make it easier to distinguish expected outliers from true errors we will accelerate troubleshooting."

Point 9

06 Provide a single explanation for all audiences

07 Undervalue explanation friendly features

08 Fail to design for debugging

09 Assume rather than demonstrate generalizability

10 Think moonshots are the model

- "We tend to be overly optimistic about our models ability to generalize. Models that work well on the lab bench and in initial production use may still have latent limitations on their ability to succeed over time, geography and use case. XAI approaches can reveal latent issues and allow non-data scientists to build confidence in how broadly a model can be applied."

Point 10

06 Provide a single explanation for all audiences

07 Undervalue explanation friendly features

08 Fail to design for debugging

09 Assume rather than demonstrate generalizability

10 Think moonshots are the model

- "When we think of AI what comes to mind first are the moonshots: the decade long projects that use breakthrough new technology to implement radical solutions to huge problems. However, consider the distribution of all machine learning projects over the next decade, only a tiny sliver of them will be these ultra expensive ground breaking projects. We need to design our common toolsets and methodologies for the median machine learning project not the moonshots."

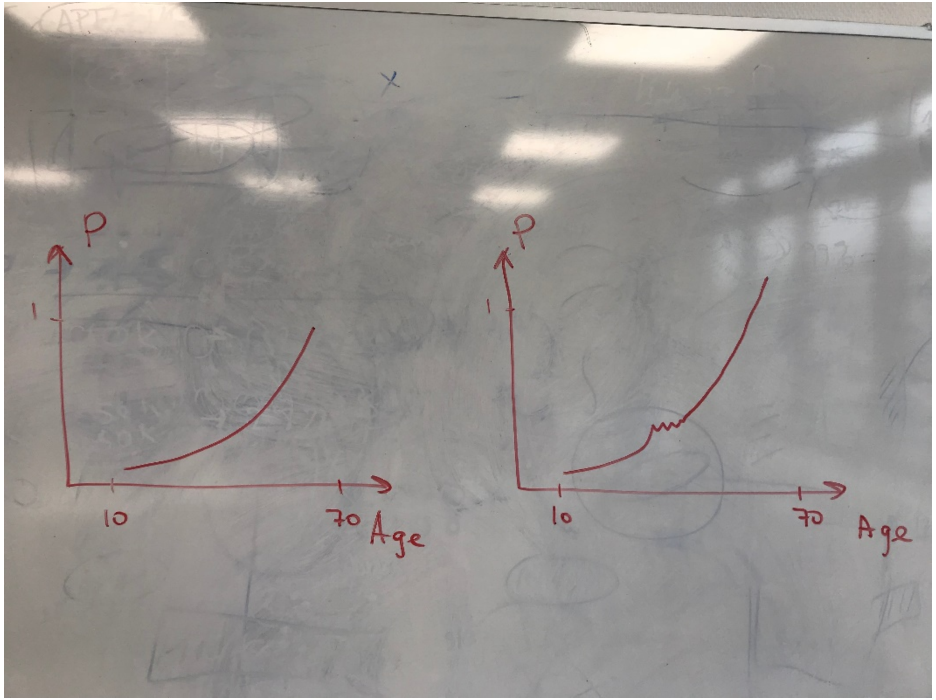
All the attention was here.

- **Is it worth spending thousands of dollars to improve prediction accuracy by 0.001% ?**

All the attention was here.

- Is it worth spending thousands of dollars to improve prediction accuracy by 0.001% ?
- When real business problems will round it all up.

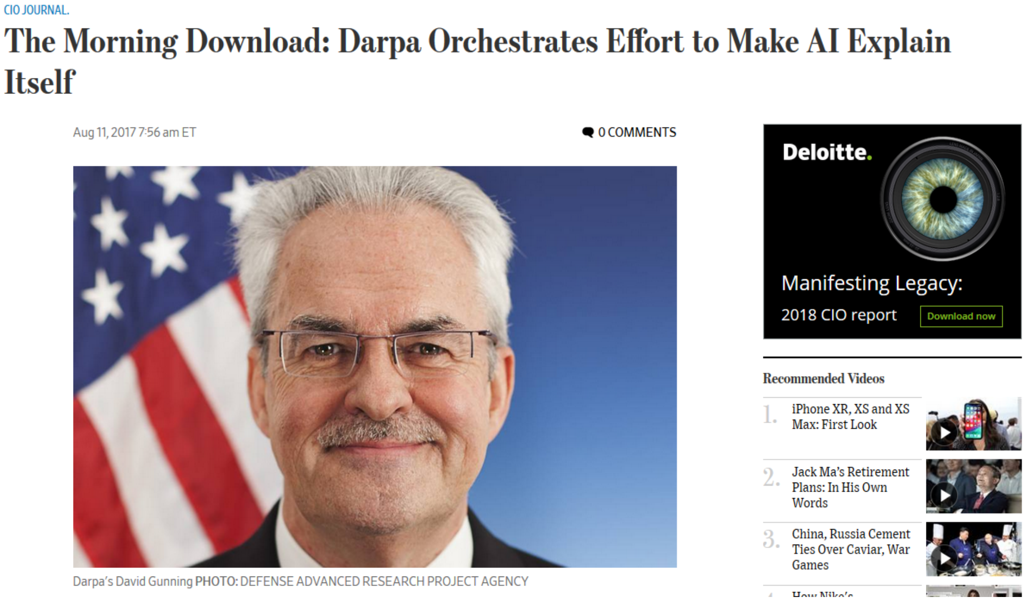
A possible situation



- Two models.
  - One with feature set A
  - The other one with an extended featue set

DARPA program

XAI will enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.



CIO JOURNAL.

The Morning Download: Darpa Orchestrates Effort to Make AI Explain Itself

Aug 11, 2017 7:56 am ET

0 COMMENTS

Darpa's David Gunning PHOTO: DEFENSE ADVANCED RESEARCH PROJECT AGENCY

Recommended Videos

1. iPhone XR, XS and XS Max: First Look
2. Jack Ma's Retirement Plans: In His Own Words
3. China, Russia Cement Ties Over Caviar, War Games

GDPR

- ✓ **GDPR Article 22 Paragraph 3 states that a data controller "shall implement suitable measures to safeguard…at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision", otherwise a person has "the right not to be subject to a decision based solely on automated processing" (Paragraph 1).**

GDPR

- ✓ GDPR Article 22 Paragraph 3 states that a data controller “shall implement suitable measures to safeguard…at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision”, otherwise a person has “the right not to be subject to a decision based solely on automated processing” (Paragraph 1).

- ✓ There are different interpretations of GDPR „right for explanation“ .

Why we need XAI?

01 Safety. We should build systems that make sound decisions.

02 Debugging. We should understand why a system does not work and how to fix it.

03 Science. We want to understand something new.

04 Mismatched Objectives. The system may not be optimized for the true objective.

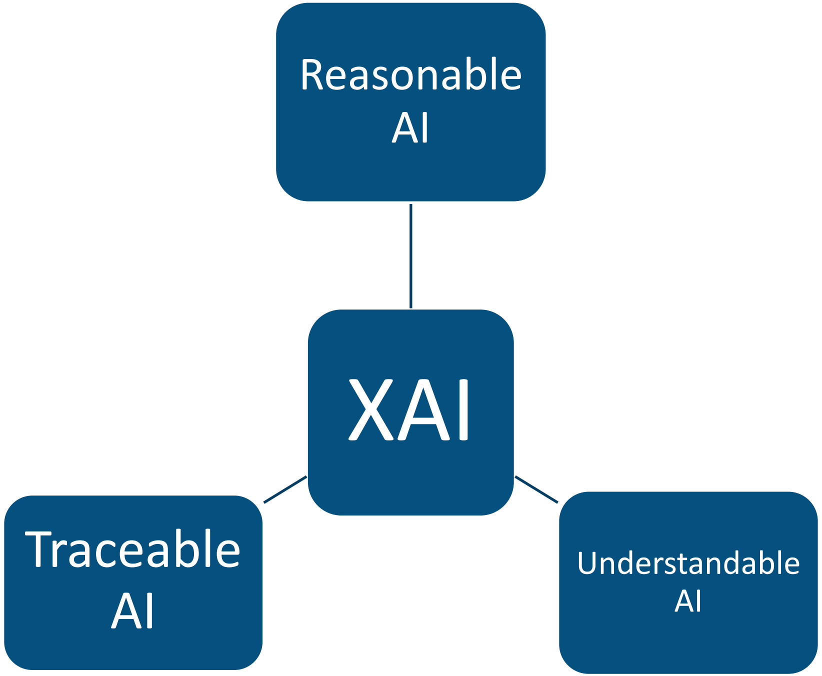
05 Legal. Are we legally required to provided an explanation?

Source https://people.csail.mit.edu/beenkim/papers/BeenK_FinaleDV_ICML2017_tutorial.pdf

In other words

Optimize	• Model performance • Decision making
Retain	• Control • Safety
Maintain	• Trust • Ethics
Comply	• Accountability • Regulation

Key features



What are good explanations?

01 Contrastive explanations. „Why P rather than Q?“

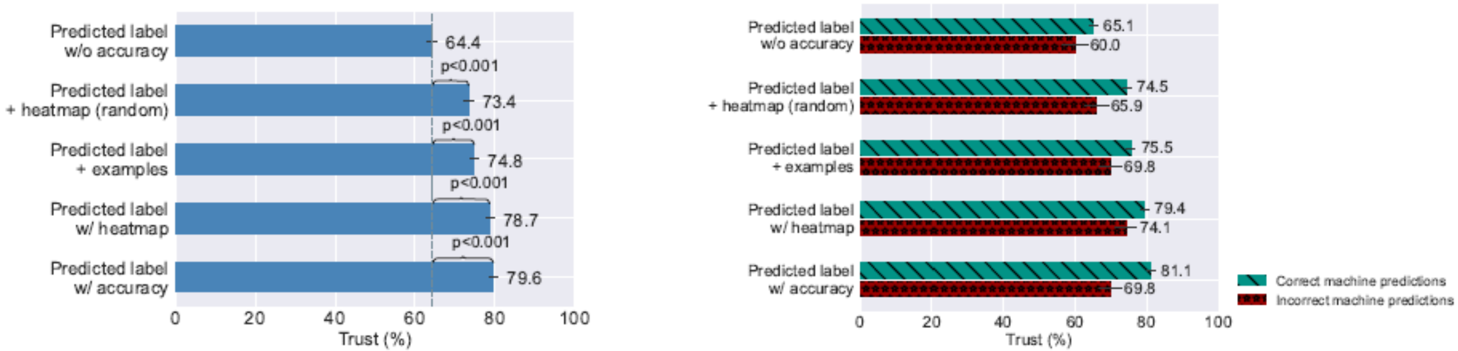
02 Social attribution.

03 Causal connection

04 Explanation selection

05 Simple, general, and more coherent.

Explanations increase trust



(a) Trust in machine predictions.

(b) Trust in correct and incorrect machine predictions.

Figure 4: The trust that humans place on machine predictions. Figure 4a shows that adding feature-based explanations (heatmap) can effectively increase the trust level compared to ***predicted label w/o accuracy***. *p*-value in Figure 4a is computed by conducting t-test between the corresponding setup and ***predicted label w/o accuracy***. Figure 4b breaks down the trust based on whether machine predictions are correct or incorrect and show that humans trust correct machine predictions more than the incorrect ones in all the five experimental setups, although the differences are only statistically significant in two setups.

What was explained

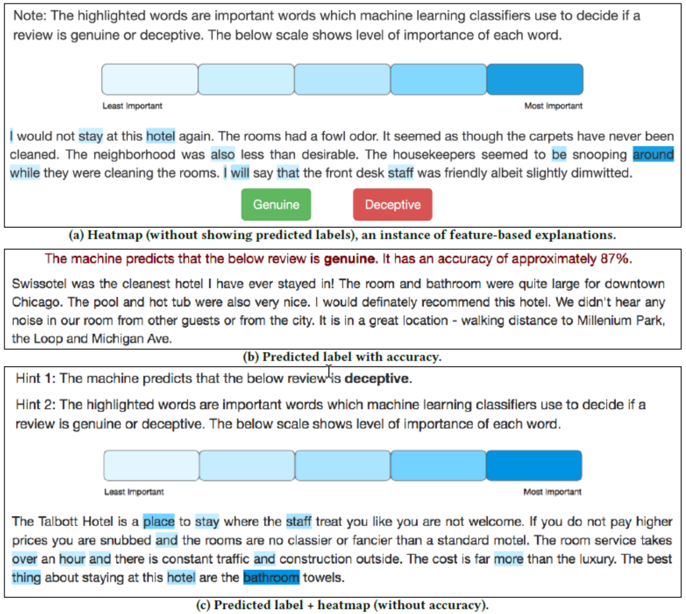


Figure 2: Example interfaces with varying levels of machine assistance. Figure 2a only presents feature-based explanations of machine predictions in the form of *heatmap*. Figure 2b shows both the predicted label and an explicit statement about machine accuracy (87%). Figure 2c shows the predicted label with heatmap, but does not present machine accuracy. We crop the "Genuine" and "Deceptive" buttons in Figure 2b and 2c to save space.

$$R_j = \sum_k \frac{x_j w_{j,k}}{\sum_j x_j w_{j,k} + \epsilon} R_k$$

Don't be scared — this equation is just weighting relevances based on neuron activation and weight connection

XAI methods.

Overview of methods and software.

01 **Explainable models**

02 Prediction explanation

- **Linear regression.**
  - Numerical feature: For an increase of the numerical feature x_j by one unit, the estimated outcome changes by β_j. An example of a numerical feature is the size of a house.
  - Extension: lasso, GLM, GAM
  - Represent only linear relationship.
  - Software: widely available.

Overview of methods and software.

01 **Explainable models**

02 Prediction explanation

- **Logistic regression.**
  - Numerical feature: For an increase of one unit of the feature x_j, the estimated odds change (multiplicatively) by a factor of $\exp(\beta_j)$.
  - Software: widely available.

Overview of methods and software.

01 **Explainable models**

02 Prediction explanation

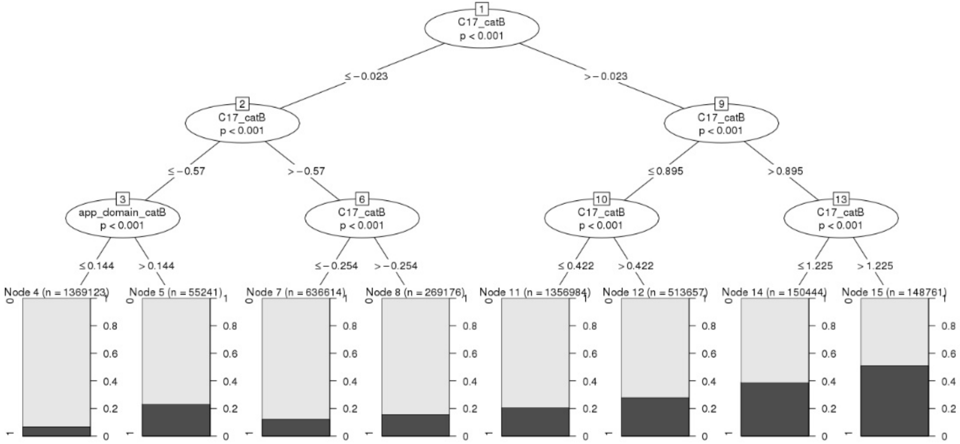
- **Decision trees.**
  - DTs split the data set into regions and provide explanations about the logic.
  - There are a lot of methods:
    - CART
    - ID3
    - Conditional Inference Trees
    - Etc.
  - Software:
    - ctree, rpart, RWeka (R)
    - sklearn (Python)
  - Advantages : cover interactions, natural visualization, good explanation.
  - Disadvantages: lack of smoothness, unstable.

Overview of methods and software.

01 **Explainable models**

02 Prediction explanation

- **Decision trees.**



Overview of methods and software.

01 **Explainable models**

02 Prediction explanation

- **Decision rule lists and decision rule sets.**
  - Produce explainable output.
  - A lot of methods appeared in the last 5 years
    - (Scalable) Baysian Rule Lists (Rudin)
    - Falling Rule Lists (Rudin)
    - Association Rule classification: CBA, QCBA, etc.
    - Certifiably Optimal Rule Lists (Rudin)
    - Interpretable Desicion Sets (Lakkaraju)
    - Etc.
  - Software:
    - sbrl (R)
    - Skater (Python)
    - Mostly code on github

Overview of methods and software.

01 **Explainable models**

02 Prediction explanation

- **Decision rule lists and decision rule sets.**
  - Advantages.
    - High level of interpretability
    - Good performance (for certain data sets)
    - Sample output

      If (city is London AND hour is 12) then prob= 0.6

      Else if (city is Berlin AND hour 15) then prob= 0.4

      Else prob = 0.1
  - Disadvantages:
    - Data need to be discretized before using such methods.

Overview of methods and software.

01 **Explainable models**

02 Prediction explanation

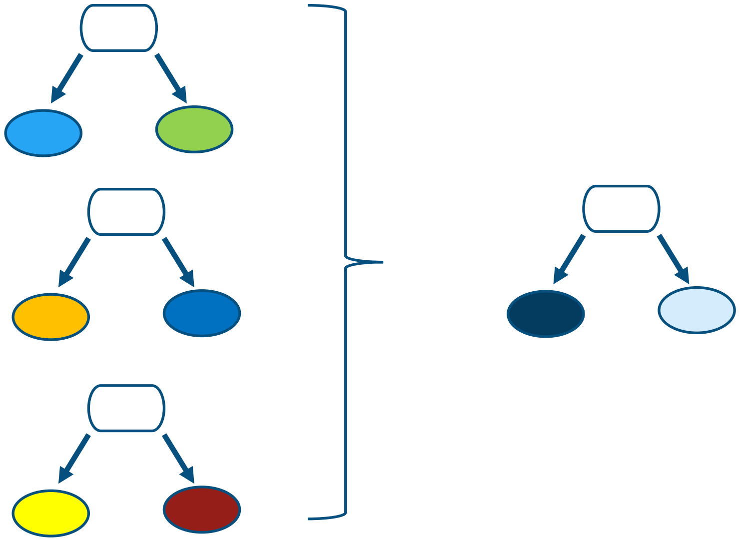
- **Ensemble to decision tree/rules**
  - Convert CART, C4.5, QUEST , GUIDE and XGBoost ensembles into a single interpretable decision tree.
  - Methods:
    - GENESIM
    - inTrees
    - ISM
    - defragTrees
  - Software
    - Github source. Try at your own risk.

Overview of methods and software.

01 Explainable models

02 Prediction explanation

- Ensemble to decision tree/rules



Overview of methods and software.

01 Explainable models

02 Prediction explanation

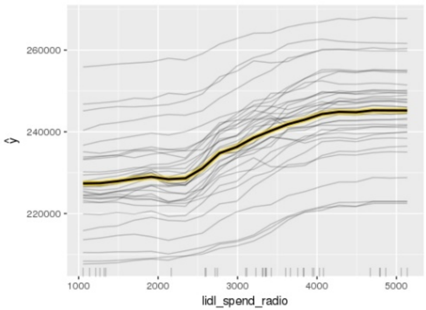
- **Visual explanations**
  - Partial dependency plot
    - Shows the marginal effect of feature on the predicted outcome of the model
    - A PDP can show if the relationship between the target and feature is linear, monotonic or complex
  - Software
    - Iml, pdp (R)
    - Skater, eli5 (python)
- **Individual Conditional expectation**
  - One line per instance, representing how the instance prediction changes when the feature changes

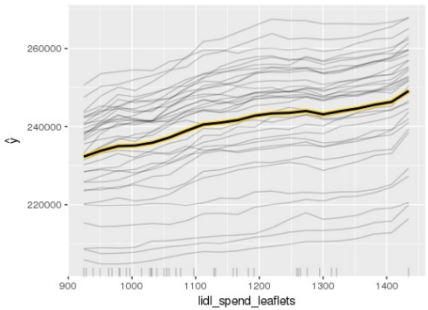
Overview of methods and software.

01 Explainable models

02 Prediction explanation

- **Visual explanations**
  - Model agnotic feature importance
    - not only a random forest can have feature importance
  - Software
    - iml, pdp (R)
    - Skater, eli5 (Python)



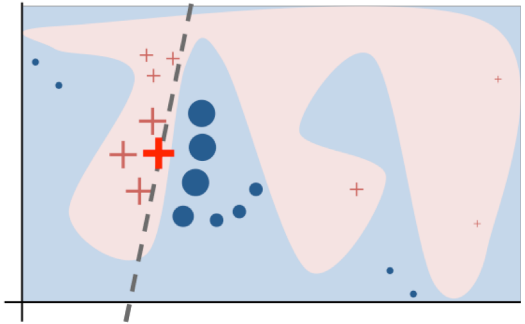


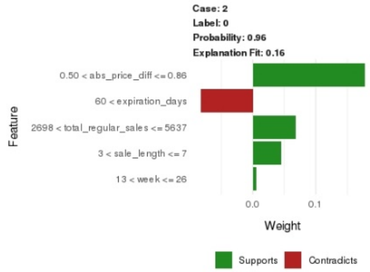
Overview of methods and software.

01 Explainable models

02 **Prediction explanation**

- **LIME**
  - It is a local surrogate model
  - It works for tabular, text, and image data
  - Software
    - Iml, lime, DALEX (R)
    - Skater, eli5 (Python)





Overview of methods and software.

01 Explainable models

02 Prediction explanation

- **LIME**
  - Disadvantages
    - Only linear model is used for local explanation.
    - Inability to know how widely one can apply a „local“ explanation
    - Dependence on a concept of closeness/distance that was vague and ill-defined from end users point of view

Overview of methods and software.

01 Explainable models

02 Prediction explanation

- **SHAP**
  - The interpretation of the Shapley value φ_{ij} for feature j and instance i is: the feature value x_{ij} contributed φ_{ij} towards the prediction for instance i compared to the average prediction for the dataset.
  - Advantages:
    - It allows for contrastive explanation
    - solid theory
  - Disadvantages:
    - expensive to compute (only approximations work)
    - need access to the original data
    - not sparse
  - Software:
    - iml, DALEX (R)
    - Skater, eli5, shap

Overview of methods and software.

01 Explainable models

02 **Prediction explanation**

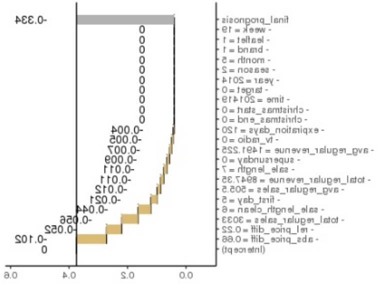
- **Breakdown**
  - The main goal is to decompose model predictions into parts that can be attributed to particular variables

$$f(x^{new}) = (1, x^{new})(\mu, \beta)^T = baseline + (x_1^{new} - \bar{x}_1)\beta_1 + \ldots + (x_p^{new} - \bar{x}_p)\beta_p$$

where

$$baseline = \mu + \bar{x}_1\beta_1 + \ldots + \bar{x}_p\beta_p.$$

  - Software:
    - breakdown, DALEX (R)



Overview of methods and software.

01 Explainable models

02 Prediction explanation

- **Anchor**
  - Model agnotic explanations based on „if-then“ rules
    - An anchor explanation is a rule that sufficiently “anchors” the prediction locally
    - For instances on which anchor holds, the prediction is (almost) always the same
  - Software
    - Github project.
  - Advantages
    - Fixes LIME problems
    - Efficient
    - Flexible (tabular, text, image data)

$$\operatorname{prec}(A) = \mathbb{E}_{\mathcal{D}(z|A)}\left[\mathbb{1}_{f(x)=f(z)}\right] \qquad P\left(\operatorname{prec}(A) \geq \tau\right) \geq 1-\delta \qquad \operatorname{cov}(A) = \mathbb{E}_{\mathcal{D}(z)}[A(z)]$$

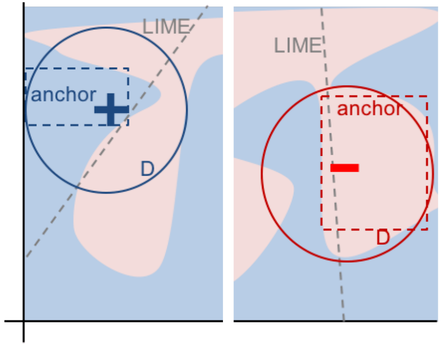
$$\max_{A \text{ s.t. } P(\operatorname{prec}(A)\geq\tau)\geq 1-\delta} \operatorname{cov}(A)$$

Overview of methods and software.

01 Explainable models

02 Prediction explanation

- **Anchor**
  - Disadvantages
    - Overly specific anchors
    - Conflicting anchors
    - Perturbation distributions

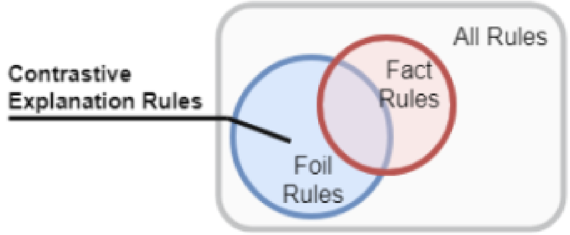


Overview of methods and software.

01 Explainable models

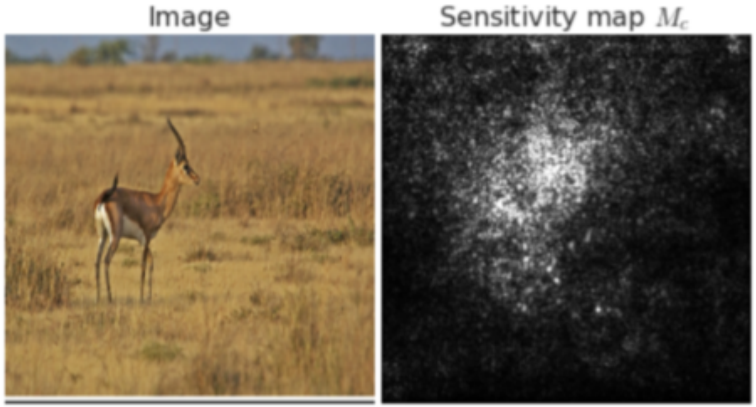
02 **Prediction explanation**

- **Contrastive explanation with foil trees.**
  - Model agnotic method
    - Answers question “why this output instead of that output”
    - Output is the form of if-then statements
  - Software
    - Github project



Methods to analyse DNNs

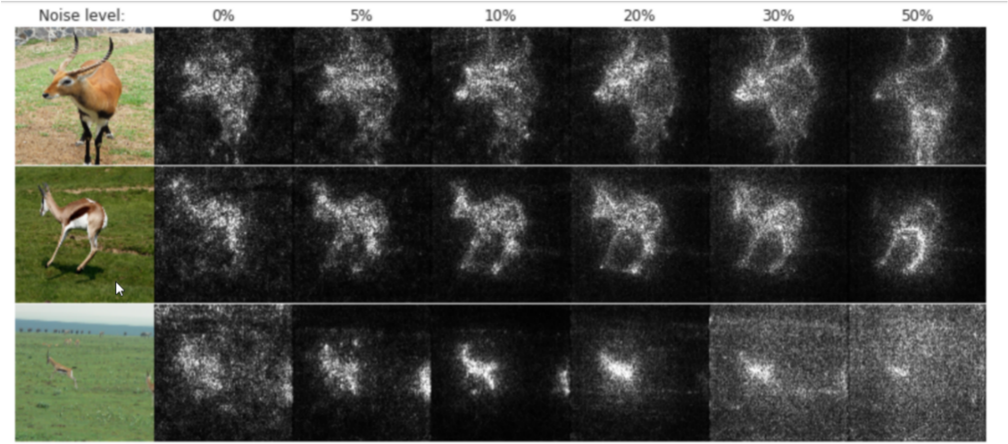
Sensitivity map based on gradient



$$class(x) = \operatorname{argmax}_{c \in C} S_c(x)$$

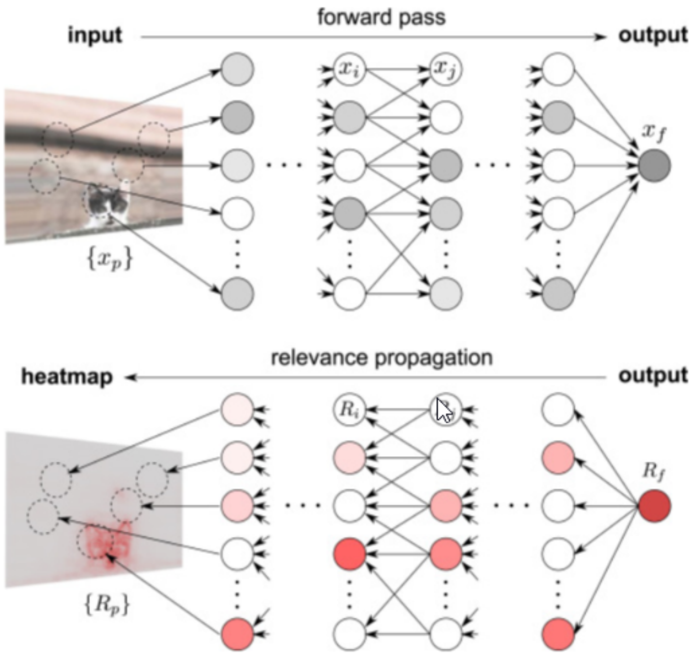
$$M_c(x) = \partial S_c(x) / \partial x$$

SmoothGrad



$$\hat{M}_c(x) = \frac{1}{n}\sum_{1}^{n} M_c(x + \mathcal{N}(0, \sigma^2))$$

Layer-wise Relevance Propagation



$$\sum_{i=1}^{d} R_i = \cdots = \sum_j R_j = \sum_k R_k = \cdots = f(x)$$

$$a_k = \sigma\left(\sum_j a_j w_{jk} + b_k\right)$$

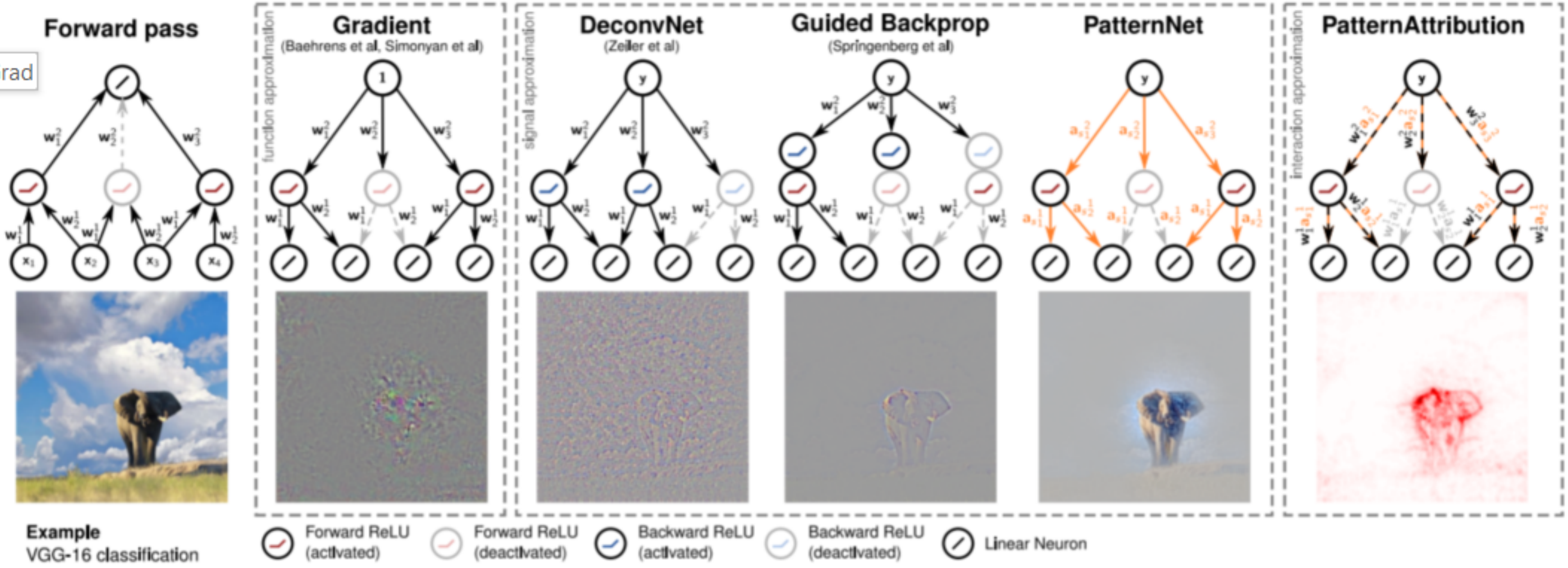
$$R_j = \sum_k \left(\alpha \frac{a_j w_{jk}^{+}}{\sum_j a_j w_{jk}^{+}} - \beta \frac{a_j w_{jk}^{-}}{\sum_j a_j w_{jk}^{-}}\right) R_k$$

http://www.heatmapping.org

Layer-wise Relevance Propagation

https://github.com/atulshanbhag/Layerwise-Relevance-Propagation

PatternNet



https://arxiv.org/abs/1705.05598

Algorithmic fairness

Fair ML.

01 **Data processing techniques.**

02 Specialized modelling techniques.

03 Predictions adjustments

04 Measuring

- Data relabelling
- Data reweighting.
- Data sampling

Source themis-ml https://arxiv.org/pdf/1710.06921.pdf

Fair ML.

01 Data processing techniques.

02 Specialized modelling techniques.

03 Predictions adjustments

04 Measuring

- Prejudice Remover Regularizer
  - Special regularization term to alleviate bias and enforce fairness
- Counterfactually fair model

Source themis-ml https://arxiv.org/pdf/1710.06921.pdf

Fair ML.

01 Data processing techniques.

02 Specialized modelling techniques.

03 **Predictions adjustments**

04 Measuring

- Reject Option Classification
  - Special procedured to assign labels near decision boundary
- Aware Ensemble Classification

Source themis-ml https://arxiv.org/pdf/1710.06921.pdf

Fair ML.

01 Data processing techniques.

02 Specialized modelling techniques.

03 Predictions adjustments

04 Measuring

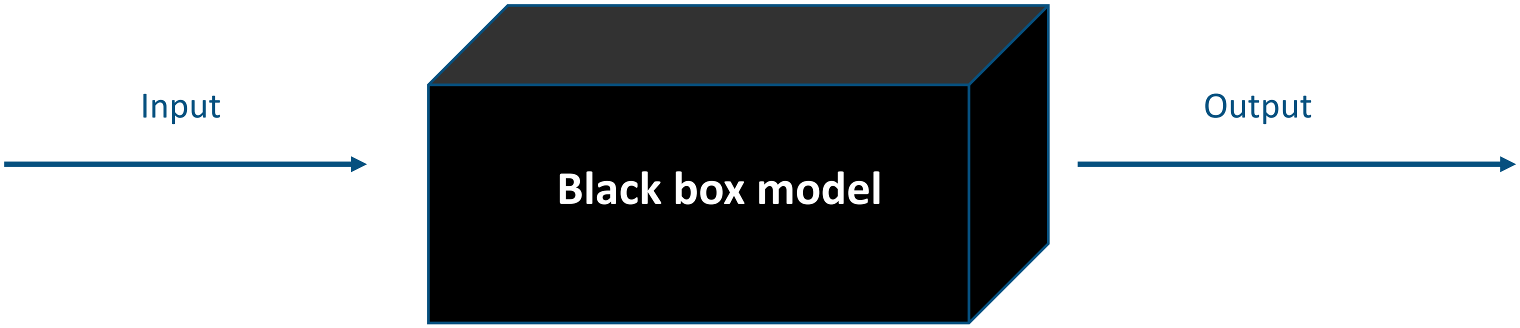
- Statistical tests
  - Difference of means test (t-test)
  - Difference of proportions for two groups (z-test)
  - Difference of proportions for many groups (Chi-Squared)
- Absolute measures
  - Mean Difference
  - Discrimination score
  - Normalized difference
  - Impact ratio
  - Odds ratio
  - Mutua information
- Conditional measures

Source themis-ml https://arxiv.org/pdf/1710.06921.pdf

XAI for programmatic advertising

What is XAI ?

Machine learning on a single slide

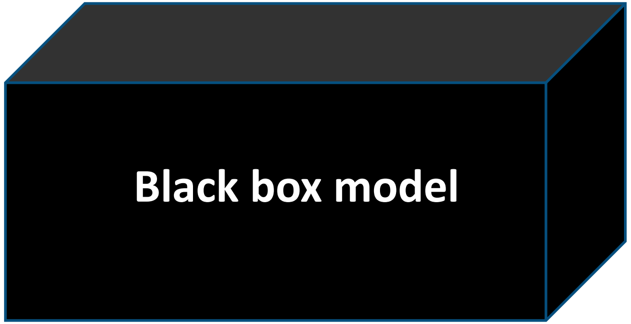


What is XAI ?

Machine learning on a single slide

Input

Click	City	Hour
1	London	12
0	London	13
0	London	14
1	Berlin	15



Output

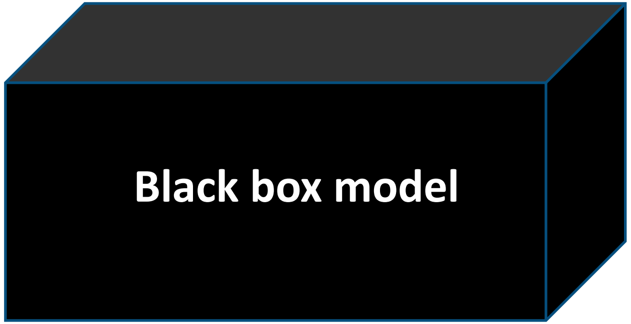
- Prediction to bid or not to bid
- Bid price prediction

What is XAI ?

Bidder optimization engine

Input

Click	City	Hour
1	London	12
0	London	13
0	London	14
1	Berlin	15



Output

- Prediction to bid or not to bid
- Bid price prediction

What is XAI ?

Bidder optimization engine

Input

Click	City	Hour
1	London	12
0	London	13
0	London	14
1	Berlin	15

$$
\begin{aligned}
(k^2 - 3K)(h_L + \tfrac{1}{3}h_T) + 3\tfrac{\dot{a}}{a}\dot{h}_L &= 4\pi G a^2[\rho_f \delta_f + \rho_s], \\
\dot{h}_L + \tfrac{1}{3}(1 - 3K/k^2)\dot{h}_T &= -4\pi G a^2[(\rho_f + p_f)v_f^{(0)} + v_s^{(0)}]/k, \\
\ddot{h}_L + \tfrac{\dot{a}}{a}\dot{h}_L &= -4\pi G a^2[\tfrac{1}{3}\rho_f \delta_f + \delta p_f + \tfrac{1}{3}\rho_s + p_s], \\
\ddot{h}_T + \tfrac{\dot{a}}{a}\dot{h}_T - k^2(h_L + \tfrac{1}{3}h_T) &= -8\pi G a^2[p_f \pi_f^{(0)} + \pi_s^{(0)}].
\end{aligned}
$$

Output

- Prediction to bid or not to bid
- Bid price prediction

What is XAI ?

Bidder optimization engine

Input

Click	City	Hour
1	London	12
0	London	13
0	London	14
1	Berlin	15

$$
\begin{aligned}
(k^2 - 3K)(h_L + \tfrac{1}{3}h_T) + 3\tfrac{\dot{a}}{a}\dot{h}_L &= 4\pi G a^2[\rho_f \delta_f + \rho_s]\,, \\
\dot{h}_L + \tfrac{1}{3}(1 - 3K/k^2)\dot{h}_T &= -4\pi G a^2[(\rho_f + p_f)v_f^{(0)} + v_s^{(0)}]/k\,, \\
\ddot{h}_L + \tfrac{\dot{a}}{a}\dot{h}_L &= -4\pi G a^2[\tfrac{1}{3}\rho_f \delta_f + \delta p_f + \tfrac{1}{3}\rho_s + p_s]\,, \\
\ddot{h}_T + \tfrac{\dot{a}}{a}\dot{h}_T - k^2(h_L + \tfrac{1}{3}h_T) &= -8\pi G a^2[p_f \pi_f^{(0)} + \pi_s^{(0)}]\,.
\end{aligned}
$$

Output

- Prediction to bid or not to bid
- Bid price prediction

How do we optimize our campaigns?

What is XAI ?

Bidder optimization engine

Input

Click	City	Hour
1	London	12
0	London	13
0	London	14
1	Berlin	15

If (city is London AND hour is 12) then Price = 2.6
Else if (city is Berlin AND hour 15) then Price = 1.4
Else Price = 1.0

Output

- Prediction to bid or not to bid
- Bid price prediction

Wouldn‘t that be nice to have?

What is XAI ?

Explainable AI

- It aims to produce "glass box" models that are explainable to a "human-in-the-loop", without greatly sacrificing AI performance.
- Human users should be able to understand the AI's cognition (both in real-time and after the fact), and should be able to determine when to trust the AI and when the AI should be distrusted

Source https://en.wikipedia.org/wiki/Marketing_mix_modeling

What is XAI and Programmatic Campaign Optimization?

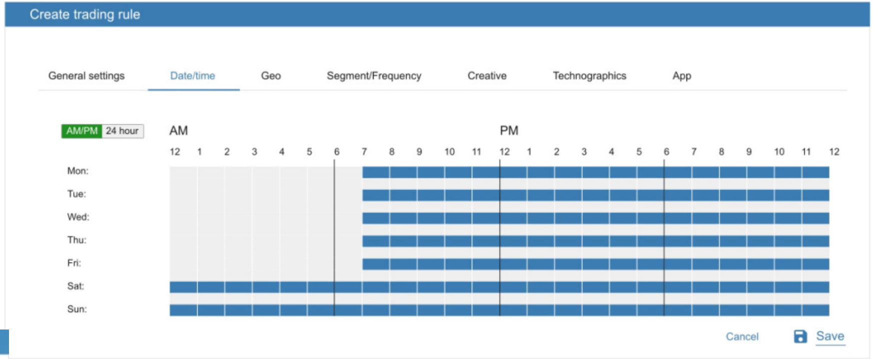
Explainable AI

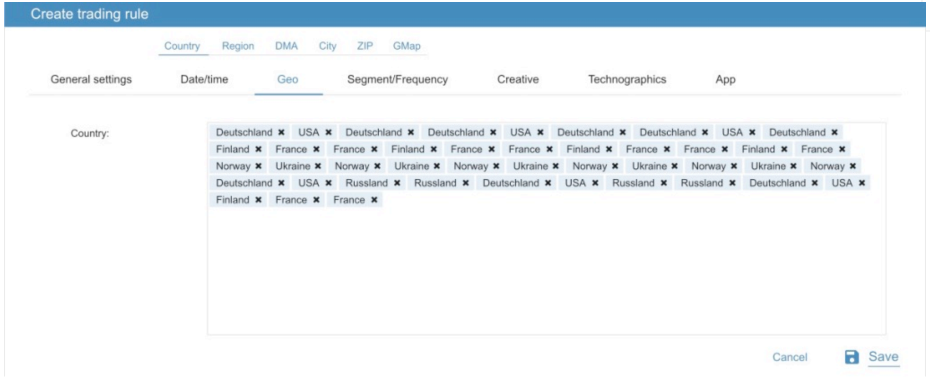
- It aims to produce "glass box" models that are explainable to a "human-in-the-loop", without greatly sacrificing AI performance.
- Human users should be able to understand the AI's cognition (both in real-time and after the fact), and should be able to determine when to trust the AI and when the AI should be distrusted

Programmatic Campaign Optimization

- One can use XAI methods to create programmatic campaigns (targeting profiles).
- One can use XAI methods to find bid prices for such campaigns.

Our programmatic campaigns. Setting hour and geo.





Methods we will explore

01 Decision trees

02 Surrogate decision trees

03 Rule lists.

Decision trees.

- Let's consider our impressions data and depict it in the form of a rectangle.

click	date	C1	banner_pos	site_category	app_domain	app_category	device_type	device_conn_type	C15	C16	C17	C18	C19	C20
0	2014-10-21	1005	0	28905ebd	7801e8d9	07d7df22	1	2	320	50	1722	0	35	-1
0	2014-10-21	1005	0	28905ebd	7801e8d9	07d7df22	1	0	320	50	1722	0	35	100084
0	2014-10-21	1005	0	28905ebd	7801e8d9	07d7df22	1	0	320	50	1722	0	35	100084
0	2014-10-21	1005	0	28905ebd	7801e8d9	07d7df22	1	0	320	50	1722	0	35	100084
0	2014-10-21	1005	1	0569f928	7801e8d9	07d7df22	1	0	320	50	2161	0	35	-1
0	2014-10-21	1005	0	f028772b	7801e8d9	07d7df22	1	0	320	50	1899	0	431	100077
0	2014-10-21	1005	0	f028772b	7801e8d9	07d7df22	1	0	320	50	2333	0	39	-1
0	2014-10-21	1005	1	f028772b	7801e8d9	07d7df22	1	0	320	50	2374	3	39	-1
1	2014-10-21	1005	0	28905ebd	7801e8d9	07d7df22	1	2	320	50	1722	0	35	-1
0	2014-10-21	1002	0	50e219e0	7801e8d9	07d7df22	0	0	320	50	2496	3	167	100191
0	2014-10-21	1005	1	f028772b	7801e8d9	07d7df22	1	0	320	50	1974	2	39	100019
0	2014-10-21	1005	0	28905ebd	7801e8d9	07d7df22	1	0	320	50	1722	0	35	-1
0	2014-10-21	1005	0	f028772b	7801e8d9	07d7df22	1	2	320	50	2161	0	35	100148
0	2014-10-21	1005	0	f028772b	7801e8d9	07d7df22	1	0	320	50	2227	0	687	100077
0	2014-10-21	1005	0	50e219e0	d9b5648e	0f2161f8	1	0	320	50	2371	0	551	-1

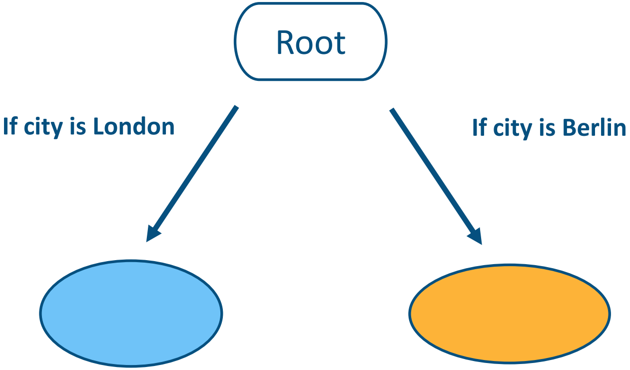
Decision trees.

- Let‘s consider our impressions data and depict it in the form of a rectangle.

Click	City	Hour
1	London	12
0	London	13
0	London	14
1	Berlin	15

London, 12	Berlin, 15
London, 12	Berlin, 15
London, 13	Berlin, 15
London, 14	Berlin, 15
London, 14	Berlin, 15

Decision trees.



Click	City	Hour
1	London	12
0	London	13
0	London	14

Click	City	Hour
1	Berlin	15

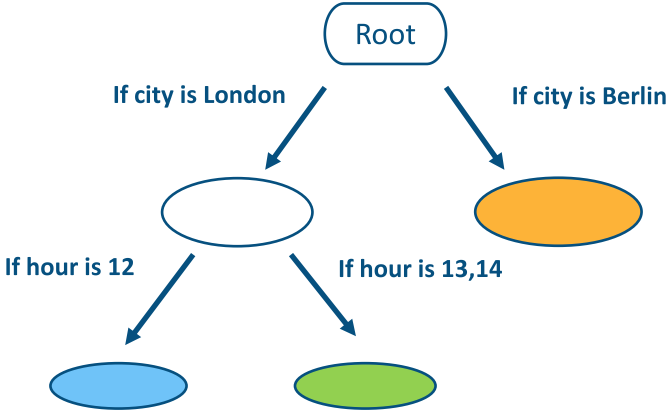
Targeting

- Geo: London
- Hour: 12, 13,14

Targeting

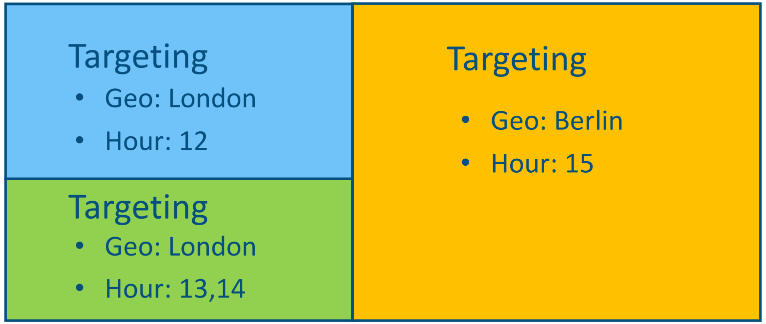
- Geo: Berlin
- Hour: 15

Decision trees.



Click	City	Hour
1	London	12
0	London	13
0	London	14

Click	City	Hour
1	Berlin	15



If (city is London AND hour is 12) then click prob. = 0.0025

if (city is Berlin AND hour 15) then click prob. = 0.0089

Else click prob. = 0.0012

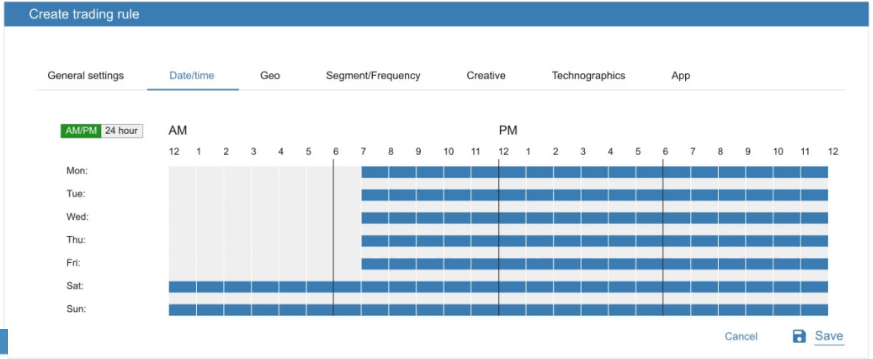
Our programmatic campaigns.

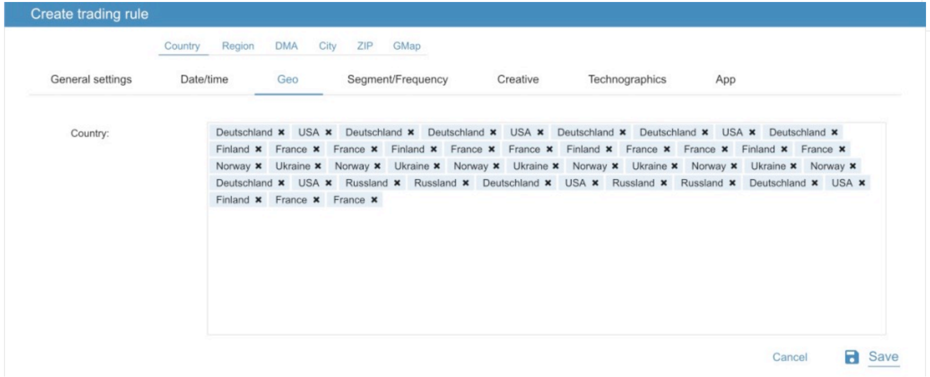
Decision tree

- splits our data into regions: blue, green, etc. Each region is a new campaign.
- has 3 leaves. Each leaf is a campaign (targeting profile).
- will output click probability for each leaf.
- Depending on the campaign goal (CPC or CPA), the bid price or CPM (cost per mille)

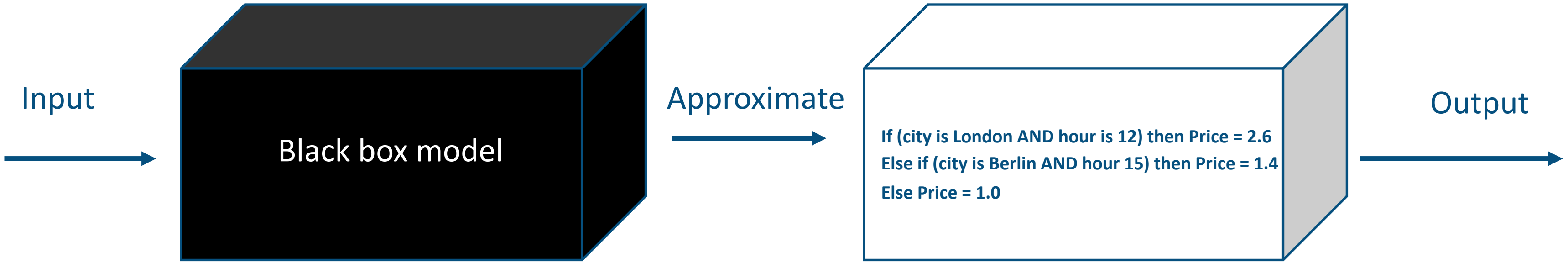
$$\text{CPM} = Goal * 1000 * Prob$$

Our programmatic campaigns. Setting hour and geo.





Surrogate decision tree.



Click	City	Hour
1	London	12
0	London	13
0	London	14
1	Berlin	15

Price	City	Hour
2.5	London	12
4.5	London	13
1.0	London	14
2.2	Berlin	15

- Prediction to bid or not to bid
- Bid price prediction

We will need predicted prices from the black box

Surrogate decision tree.

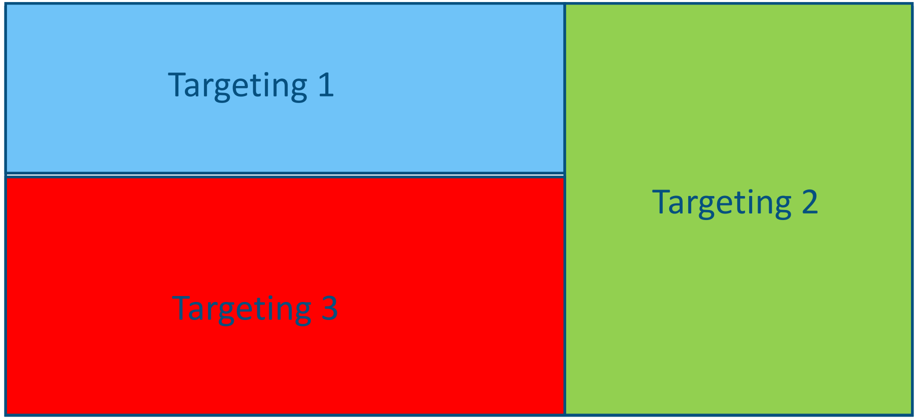
Surrogate decision tree

- will work when you already have a black box model and the original data at hand. It will generate targeting profiles and bid prices for you.
- will work when you have predictions of a black box model and the original data at hand.
- may be useful when a simple decision tree fails and would like to get an explainable model out of a black box.
- has interpretation analagous to the direct decision trees described above.

Decision rule lists.

- Let‘s consider our impressions data and depict it in the form of a rectangle.

Click	City	Hour
1	London	12
0	London	13
0	London	14
1	Berlin	15



If (city is London AND hour is 12) then click prob. = 0.0025

Else if (city is Berlin AND hour 15) then click prob. = 0.0089

Else click prob. = 0.0012

Decision rule lists

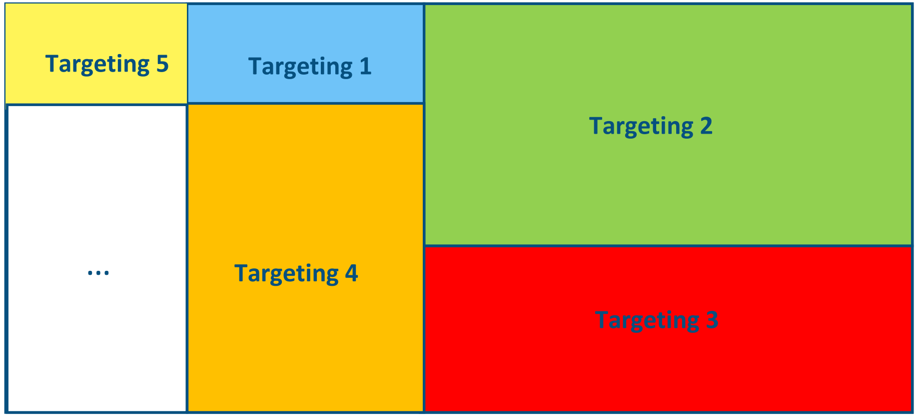
Decision rule lists model

- is the the state-of-the-art technique in the field of explainable models.
- generates a series of if-else statements that can be directly translated into targeting profiles and probabilities into bid prices

Build a strategy.

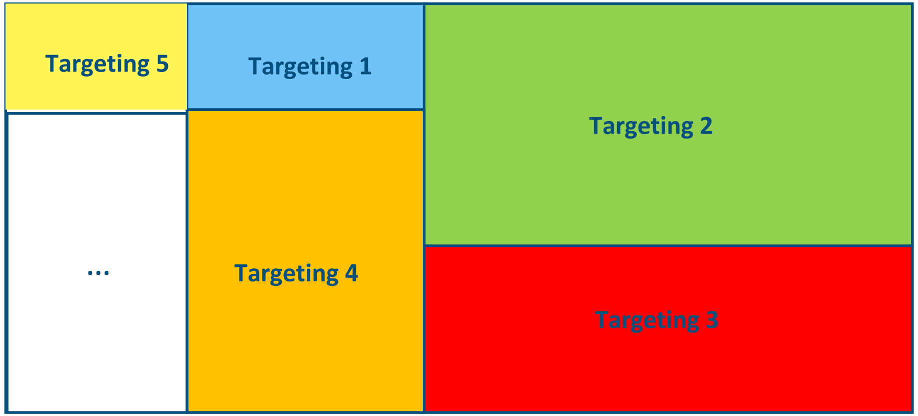
- Depending on model parameters one can generate a lot or a few campaigns.

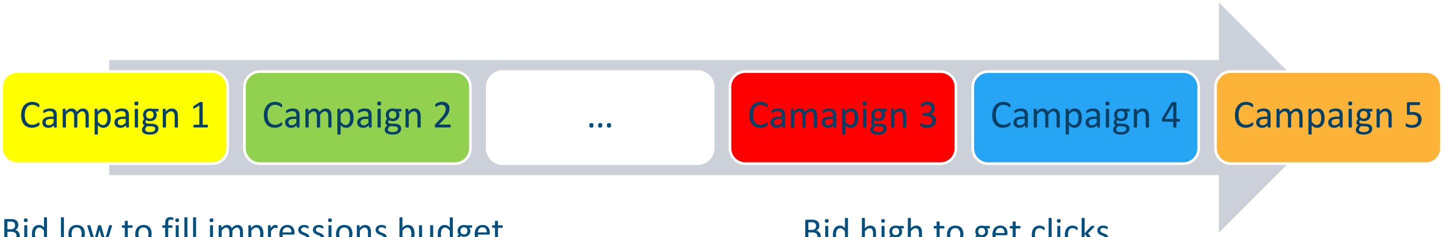
Click	City	Hour
1	London	12
0	London	13
0	London	14
1	Berlin	15



Build a strategy.

- Depending on model parameters one can generate a lot or a few campaigns.
- Sort the according to probability/bid price





Bid low to fill impressions budget

Bid high to get clicks

Summary. Steps to build campaigns.

Collect impressions log

Build a decision tree or a rule list/set

Transform tree leaves or rule conditions into targeting profiles

Compute bid price from probabilities

XAI for marketing.

Marketing Mix Model.

Marketing mix modeling (**MMM**) is statistical analysis such as multivariate regressions on sales and marketing time series data to estimate the impact of various marketing tactics (marketing mix) on sales and then forecast the impact of future sets of tactics. It is often used to optimize advertising mix and promotional tactics with respect to sales revenue or profit.

Marketing Mix Model Sample (fake) Data

Revenue	Radio spend	TV spend	Cinema spend	Outdoor spend
111111	234	125	100	111
121212	345	122	200	167
131213	134	111	139	89

- One can try to build a blackbox model in order to predict revenue based on different marketing chennels spends.

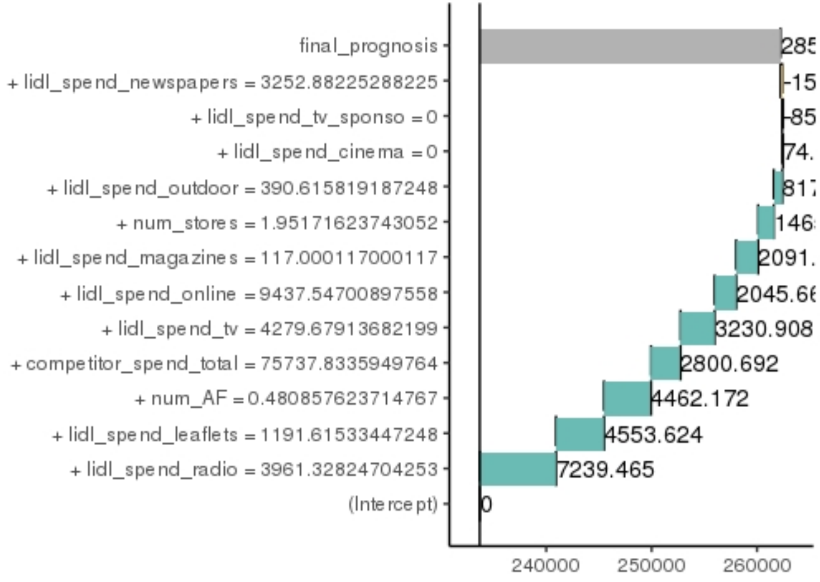
Marketing Mix Model Sample (fake) Data

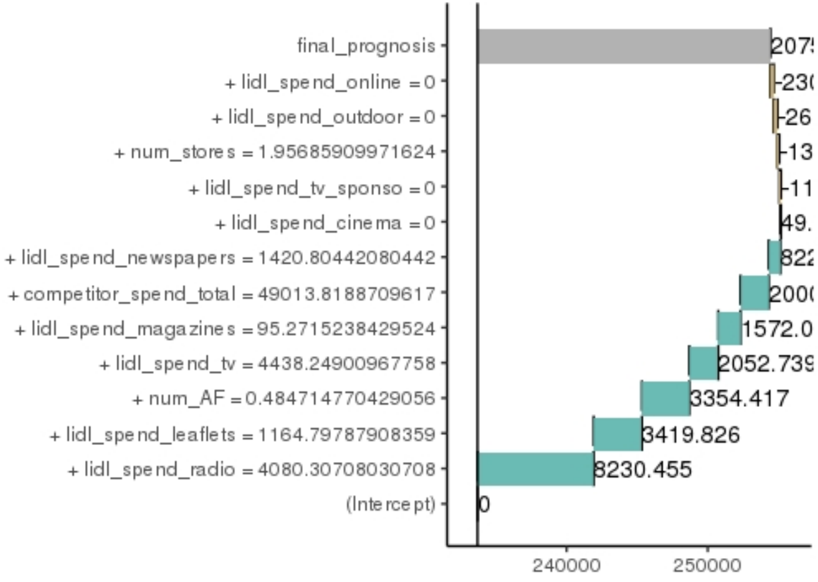
Revenue	Radio spend	TV spend	Cinema spend	Outdoor spend
111111	234	125	100	111
121212	345	122	200	167
131213	134	111	139	89

- One can try to build a blackbox model in order to predict revenue based on different marketing chennels spends.

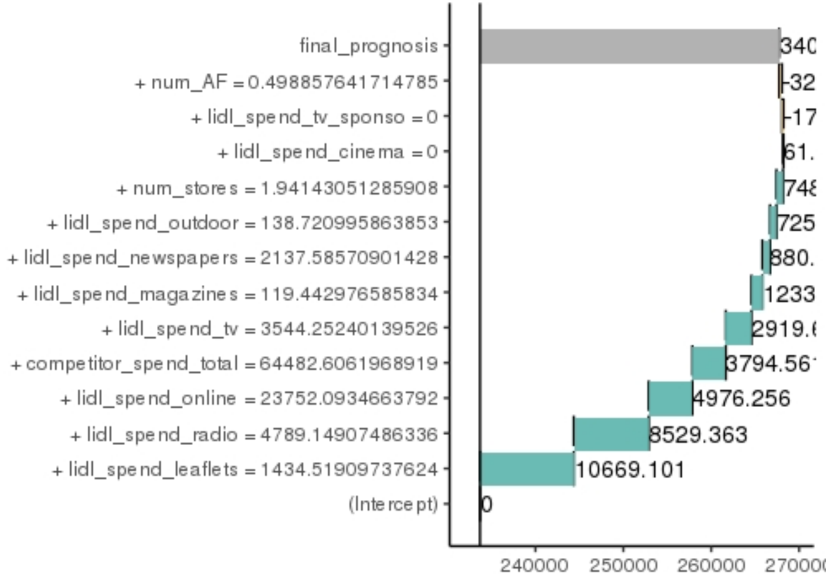
- How can one make sense out of that blackbox?

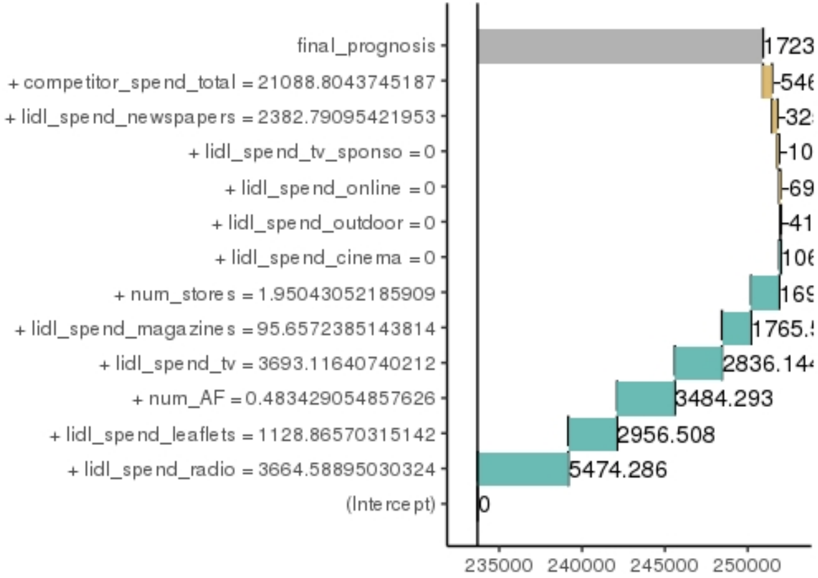
Marketing Mix Model. Channel contribution breakdown.



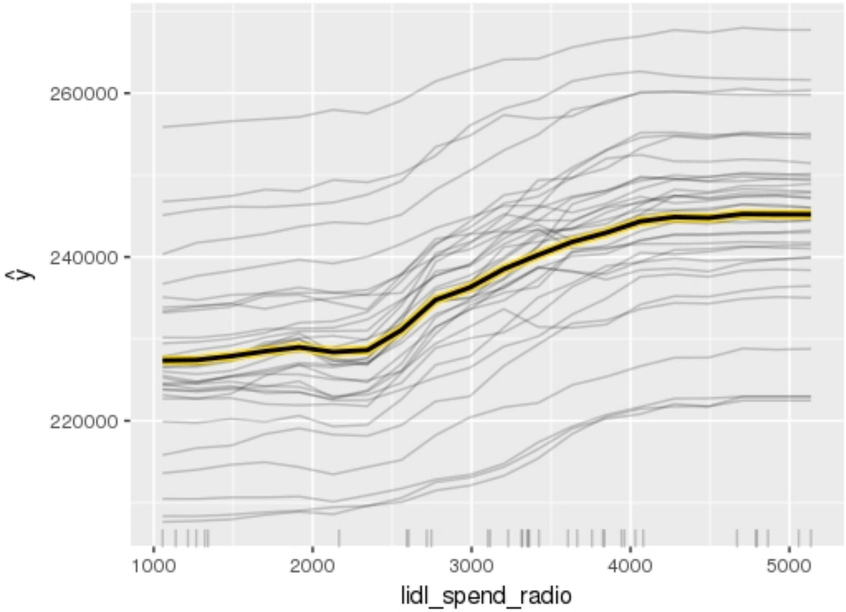


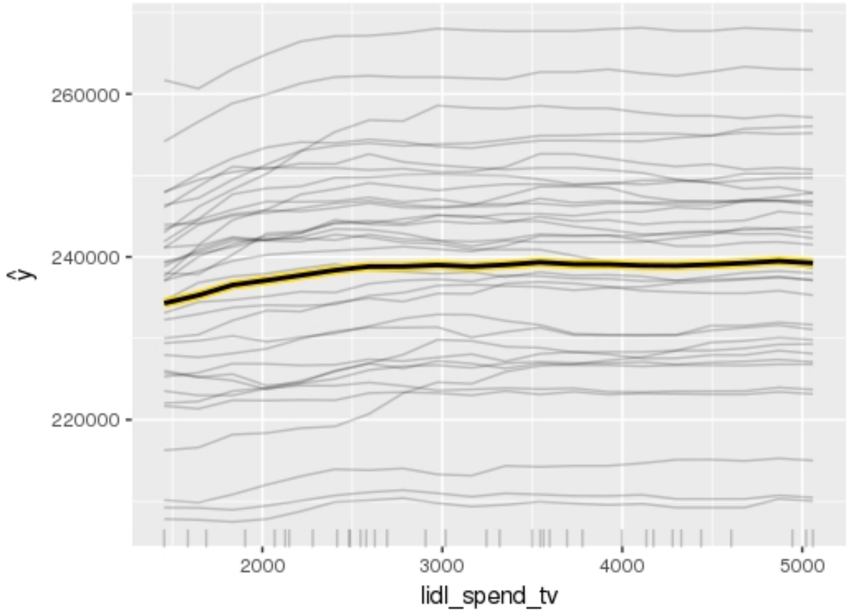
Marketing Mix Model. Channel contribution breakdown.



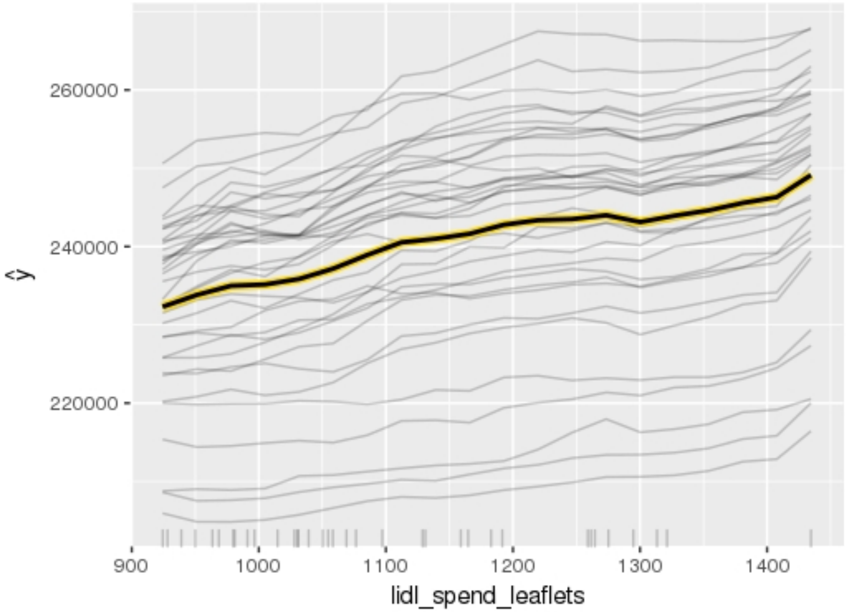


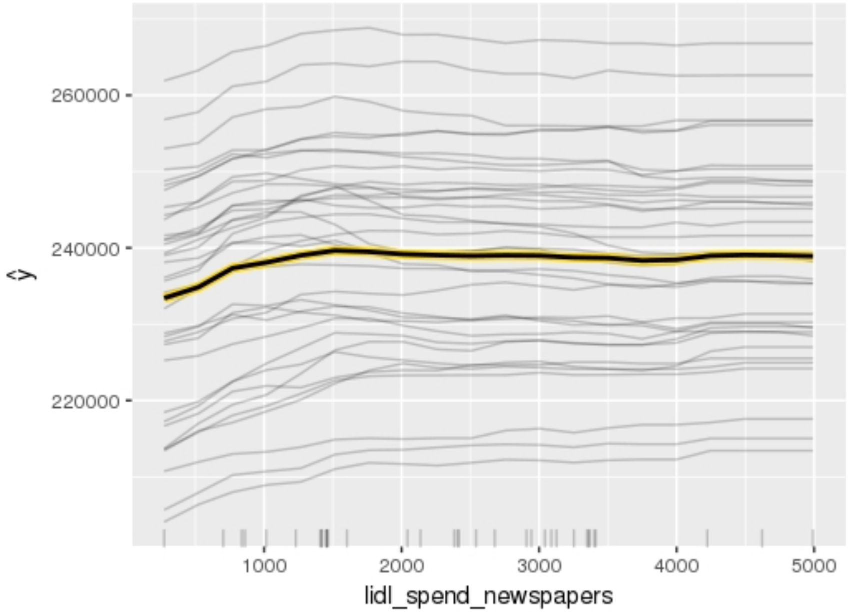
Channel contribution



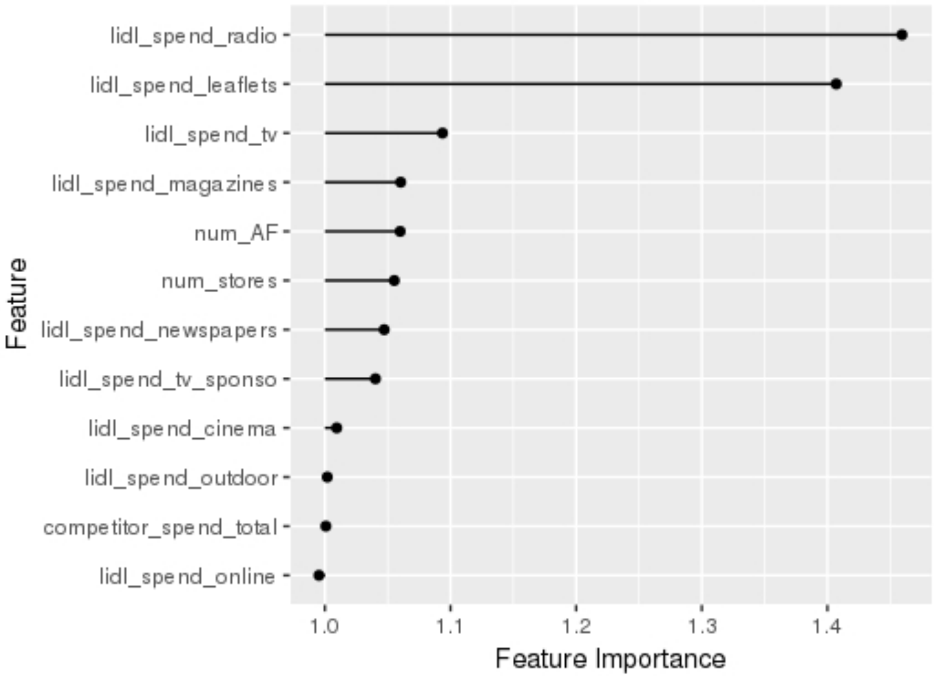


Channel contribution.

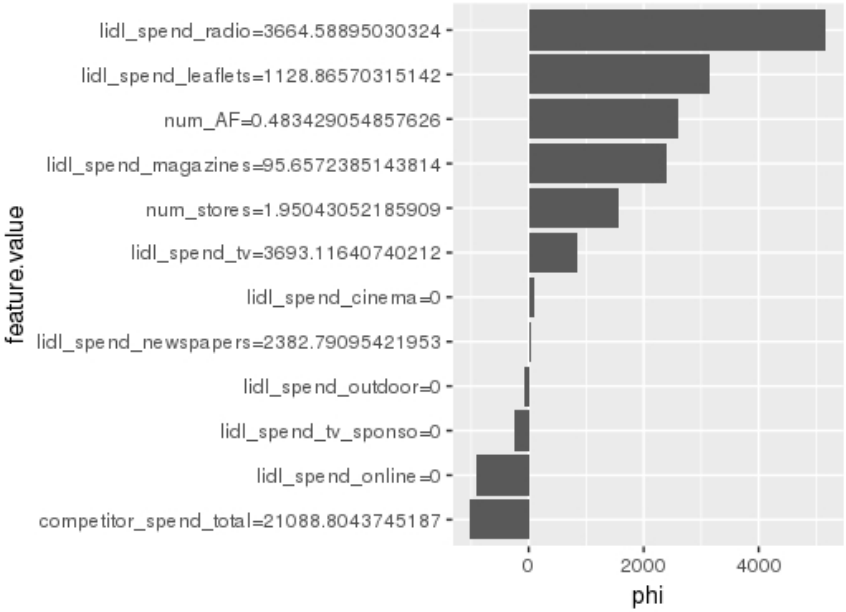


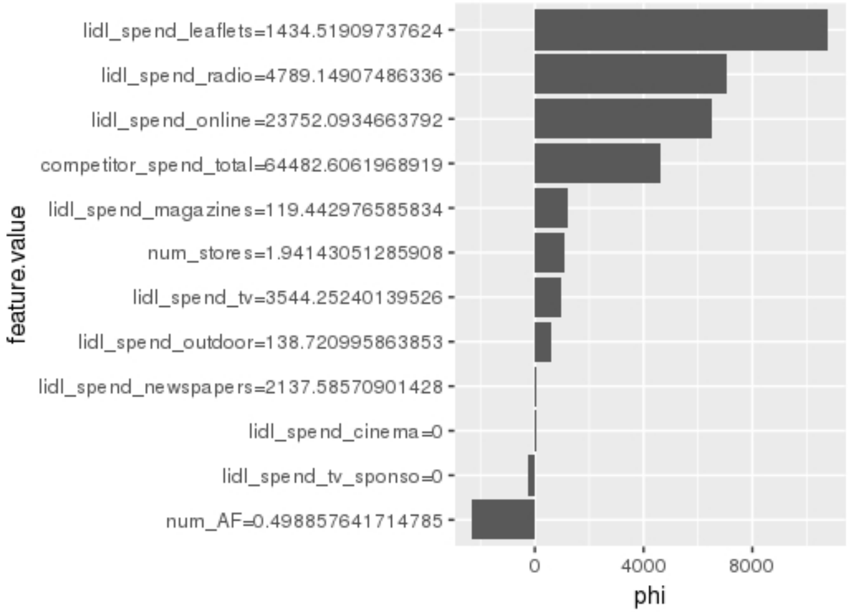


Channel importance



Channel contributions





Value of XAI for marketing.

Function	Traditional ML	XAI	XAI Value
Online Recommendations	Predicts next best recommendation	Predicts what a customer wants and why for more precise personalized recommendation messaging	More personalized recommendations reduces abandoned shopping carts, increases basket size and conversion rates
Outbound Campaigns	Automates who should get a campaign promotion	Automates and provides the contextual relevance associated with a campaign promotion for messaging purposes	Provides the drivers behind predicted behavior for greater relevance and increased sales conversion rates
Customer Analysis	Predicts customer behavior	Predicts and reveals why customers will behave as predicted	Makes predictions actionable in a real world way

Source https://simmachines.com/explainable-ai/

XAI in retail

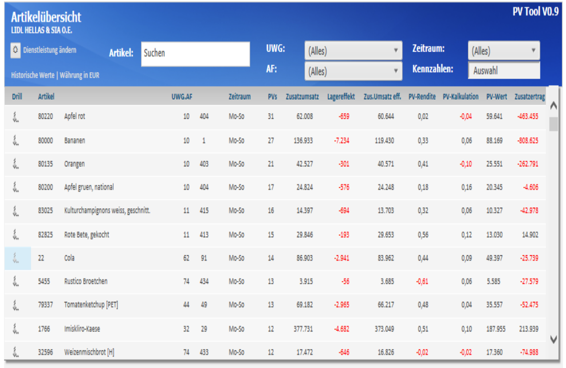
PV-Tool

Initial State

- Numerous PV campaigns in the individual Lidl countries with z.T. different requirements and framework conditions
- High investment in assortment PVs based on experience of the action leaders
- No nationwide data basis for planning
- No support through predictive analytics

Goals and mission

- Better data for better decisions in the PV business
- Uniform methodology, transparency and effort reduction
- Optimization and forecast of additional sales, frequency, yield and PV value
- Potentials increase through analysis of price elasticity



Artikelübersicht
LIDL HELLAS & SIA O.E.

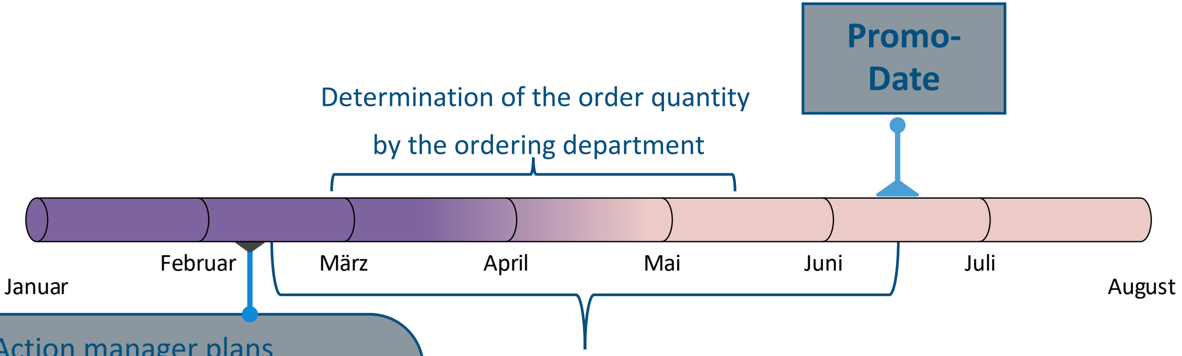
PV Tool V0.9

Dienstleistung ändern | Artikel: Suchen | UWG: (Alles) | AF: (Alles) | Zeitraum: (Alles) | Kennzahlen: Auswahl

Historische Werte | Währung in EUR

Drill	Artikel		UWG.AF		Zeitraum	PVs	Zusatzumsatz	Lagereffekt	Zus.Umsatz eff.	PV-Rendite	PV-Kalkulation	PV-Wert	Zusatzertrag
	80220	Apfel rot	10	404	Mo-So	31	62.008	-659	60.644	0,02	-0,04	59.641	-463.455
	80000	Bananen	10	1	Mo-So	27	136.933	-7.234	119.430	0,33	0,06	88.169	-808.625
	80135	Orangen	10	403	Mo-So	21	42.527	-301	40.571	0,41	-0,10	25.551	-262.791
	80200	Apfel gruen, national	10	404	Mo-So	17	24.824	-576	24.248	0,18	0,16	20.345	-4.606
	83025	Kulturchampignons weiss, geschnitt.	11	415	Mo-So	16	14.397	-694	13.703	0,32	0,06	10.327	-42.978
	82825	Rote Bete, gekocht	11	413	Mo-So	15	29.846	-193	29.653	0,56	0,12	13.030	14.902
	22	Cola	62	91	Mo-So	14	86.903	-2.941	83.962	0,44	0,09	49.397	-25.739
	5455	Rustico Broetchen	74	434	Mo-So	13	3.915	-56	3.685	-0,61	0,06	5.585	-27.579
	79337	Tomatenketchup [PET]	44	49	Mo-So	13	69.182	-2.965	66.217	0,48	0,04	35.557	-52.475
	1766	Imiskliro-Kaese	32	29	Mo-So	12	377.731	-4.682	373.049	0,51	0,10	187.955	213.939
	32596	Weizenmischbrot [H]	74	433	Mo-So	12	17.472	-646	16.826	-0,02	-0,02	17.360	-74.988

How does it work?



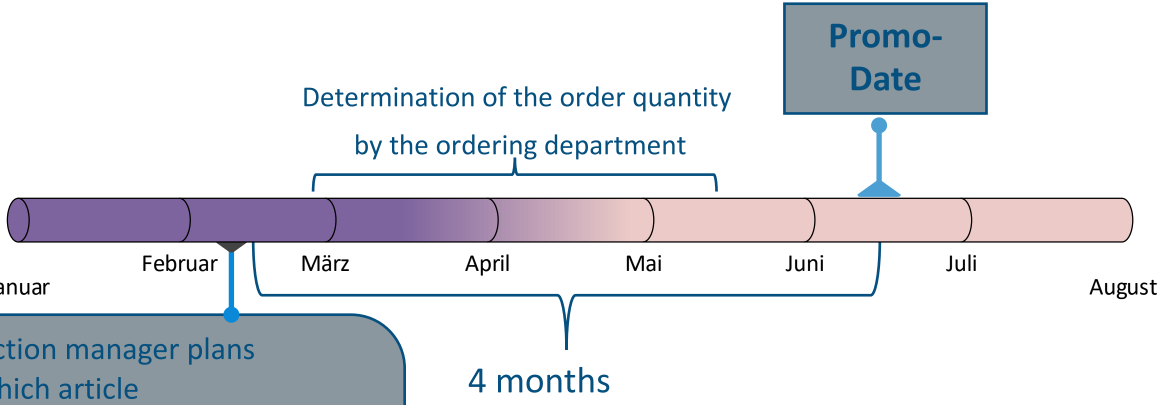
- Action manager plans
  which article
  what discount
  which marketing measure
  with which duration
  is advertised

So far, the AM has been oriented towards the planning based on historical PVs.

Challenges for the action manager:

- Find the "best" comparable action manually
- Difficult with new articles / discount levels / duration / marketing measures etc.

How does it work?



- Action manager plans
  which article
  what discount
  which marketing measure
  with which duration
  is advertised

So far, the AM has been oriented towards the planning based on historical PVs.

Challenges for the action manager:

- Find the "best" comparable action manually
- Difficult with new articles / discount levels / duration / marketing measures etc.

Let‘s predict sales uplift

ML for business. Challenges

Can you explain your model?

We have a model !

We don‘t trust models

Code

```r
fitControl <- trainControl(## structure for the CV
  verboseIter = TRUE,
  method = "repeatedcv",
  number = 5,
  repeats = 5,
  selectionFunction = "best") #alternative: "oneSE"

model <- train(x = X_train, y = y_train,
               method = "rf", preProcess = NULL,
               trControl = fitControl,
               tuneGrid = expand.grid(mtry = seq(floor(ncol(X_train)/4), floor(ncol(X_train)/1.5), 3)),
               ntree=100, keep.inbag = FALSE, importance = FALSE, replace = FALSE, maxnodes = 5)

predictor <- Predictor$new(model, data = X_test, y = y_test)
```

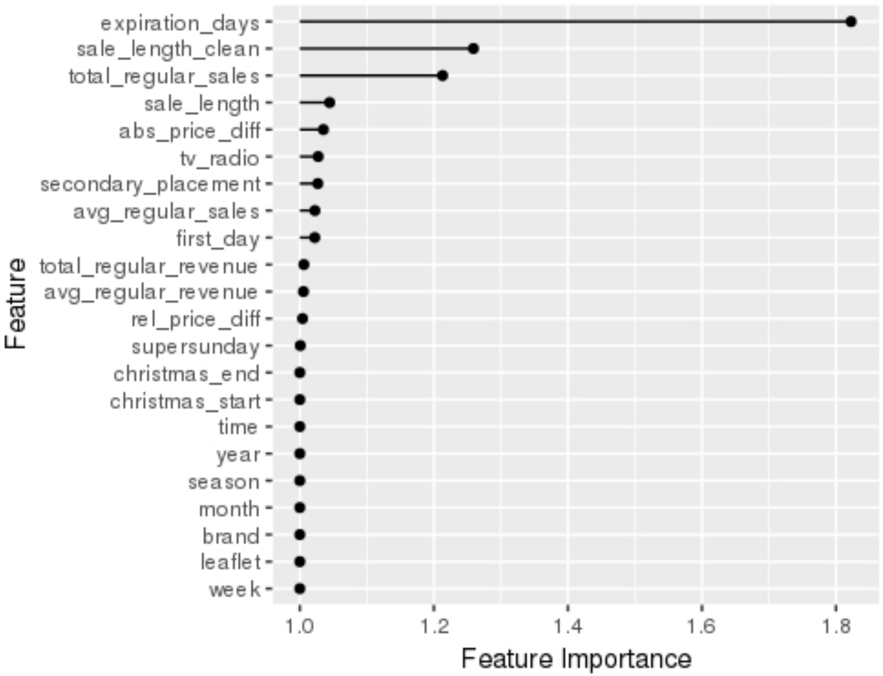
Code. PDP and Importance

```
imp = FeatureImp$new(predictor, loss = "mse")
plot(imp)

pdp.obj = Partial$new(predictor, feature = "expiration_days")
pdp.obj$plot()

pdp.obj = Partial$new(predictor, feature = "sale_length_clean")
pdp.obj$plot()

pdp.obj = Partial$new(predictor, feature = "total_regular_sales")
pdp.obj$plot()

pdp.obj = Partial$new(predictor, feature = "abs_price_diff")
pdp.obj$plot()

pdp.obj = Partial$new(predictor, feature = "supersunday")
pdp.obj$plot()
```

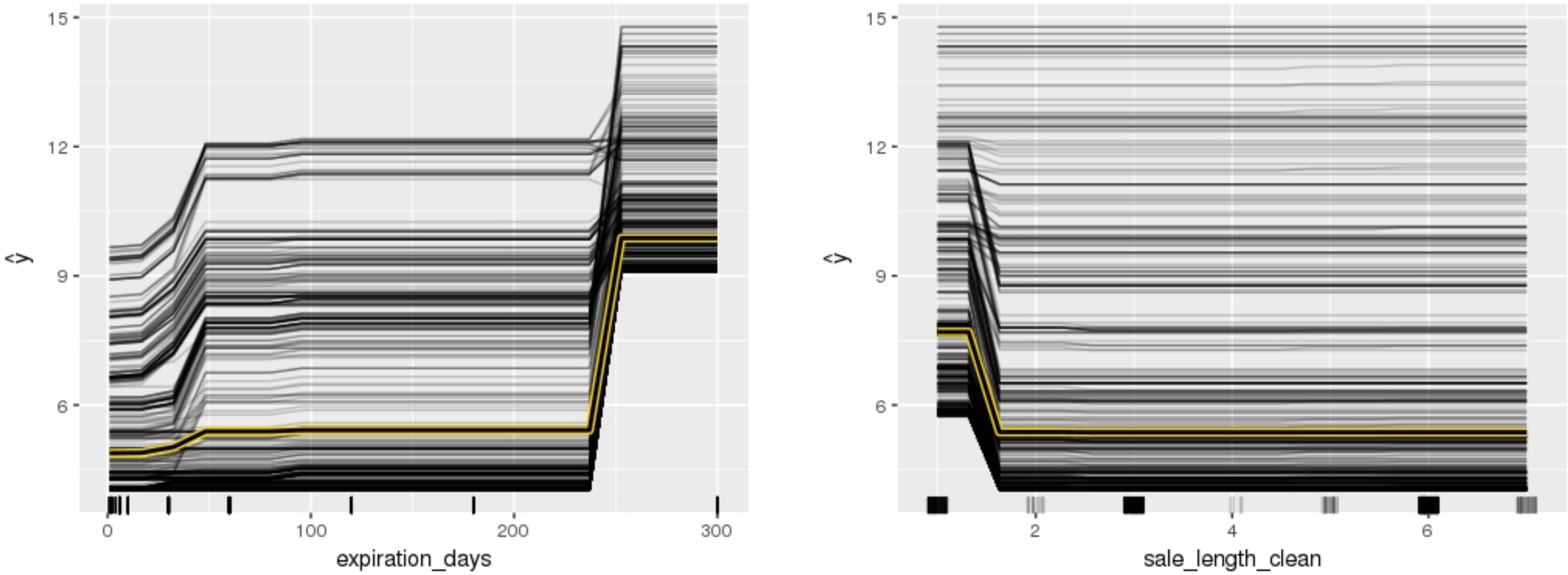
Code. LIME, SHAP, and breakDown

```
shapley = Shapley$new(predictor, x.interest = X_test[1,])
shapley$plot()

shapley = Shapley$new(predictor, x.interest = X_test[100,])
shapley$plot()

# lime package

explainer <- lime(X_train, model)
explanation <- lime::explain(X_test[1,], explainer, n_features = 5)
plot_features(explanation)

explainer <- lime(X_train, model)
explanation <- lime::explain(X_test[100,], explainer, n_features = 5)
plot_features(explanation)

# breakDown

explain_br <- broken(model, new_observation = df_test[1,], data = df_train, baseline = "intercept", direction = "up")
plot(explain_br)

explain_br <- broken(model, new_observation = df_test[100,], data = df_train, baseline = "intercept", direction = "up")
plot(explain_br)
```

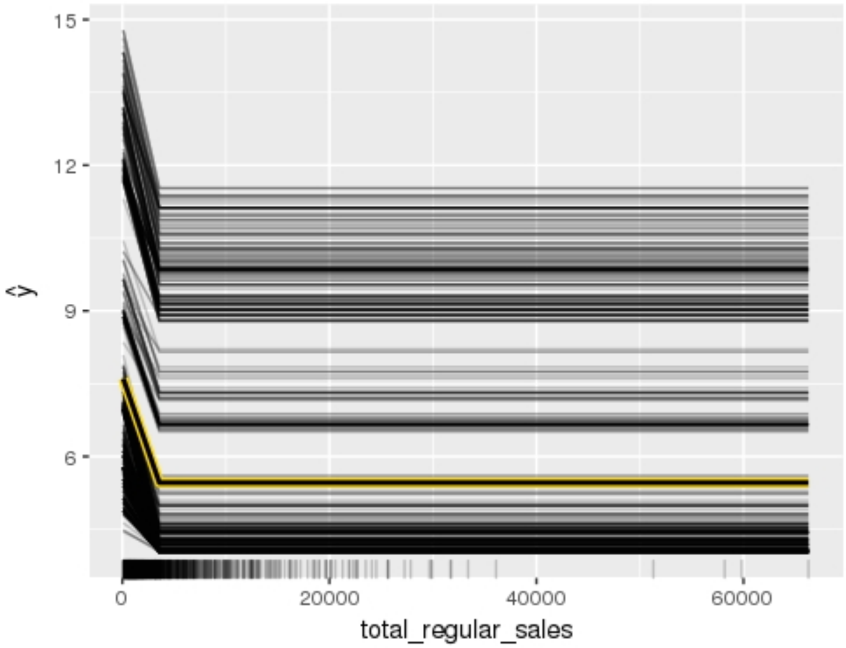
Regression. Variable importance.

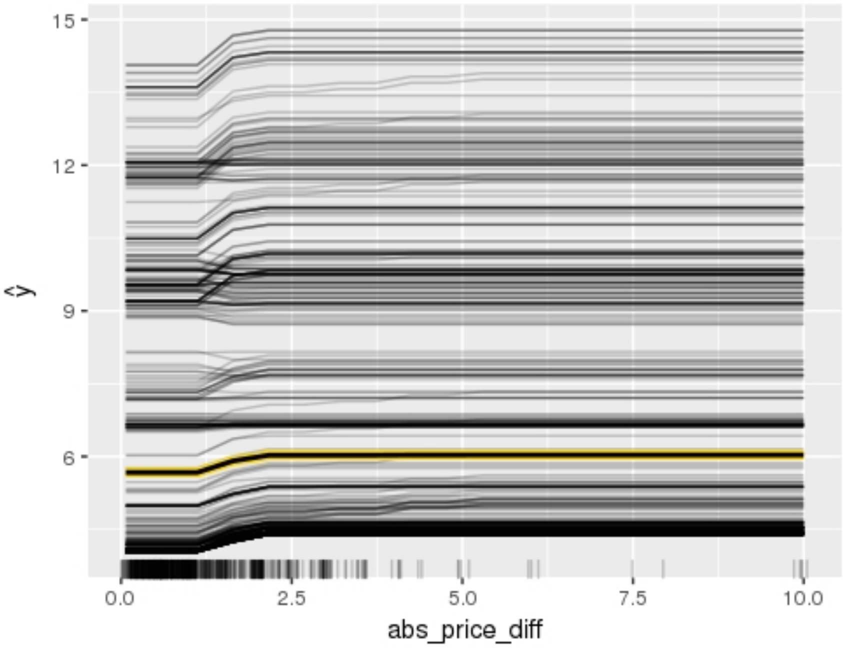


Regression. PDP.

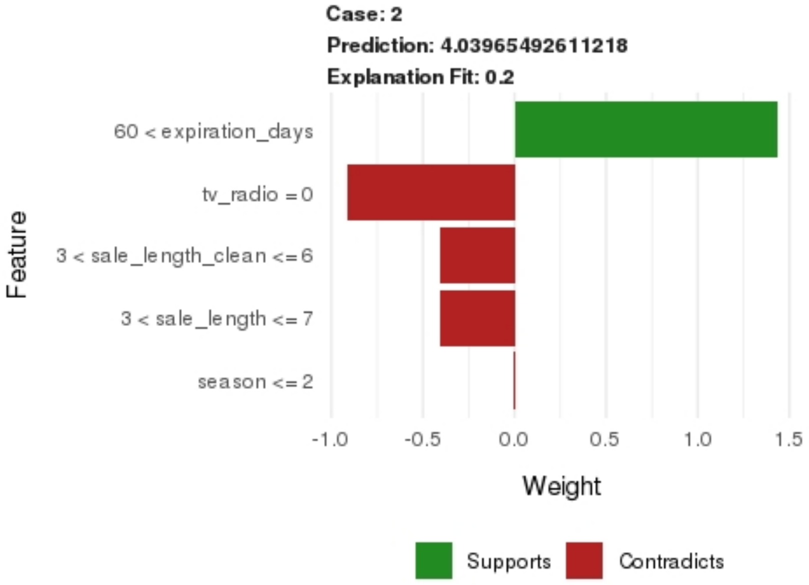


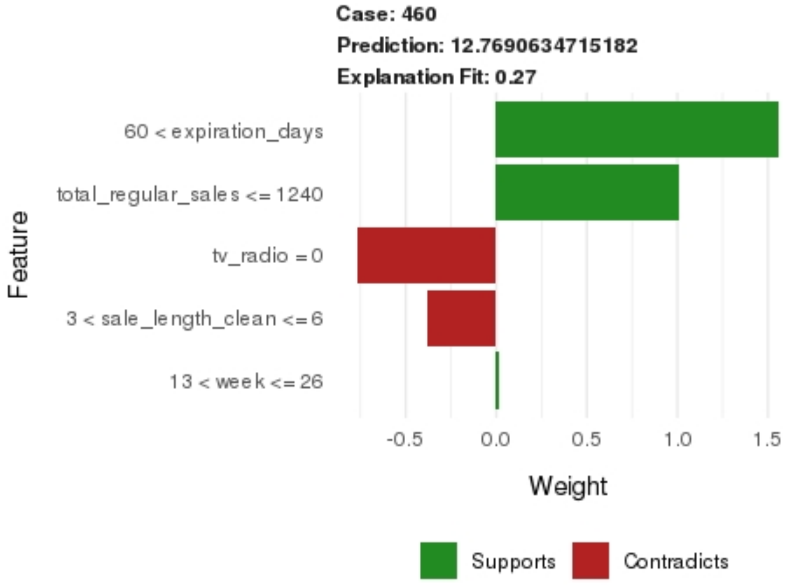
Regression. PDP.



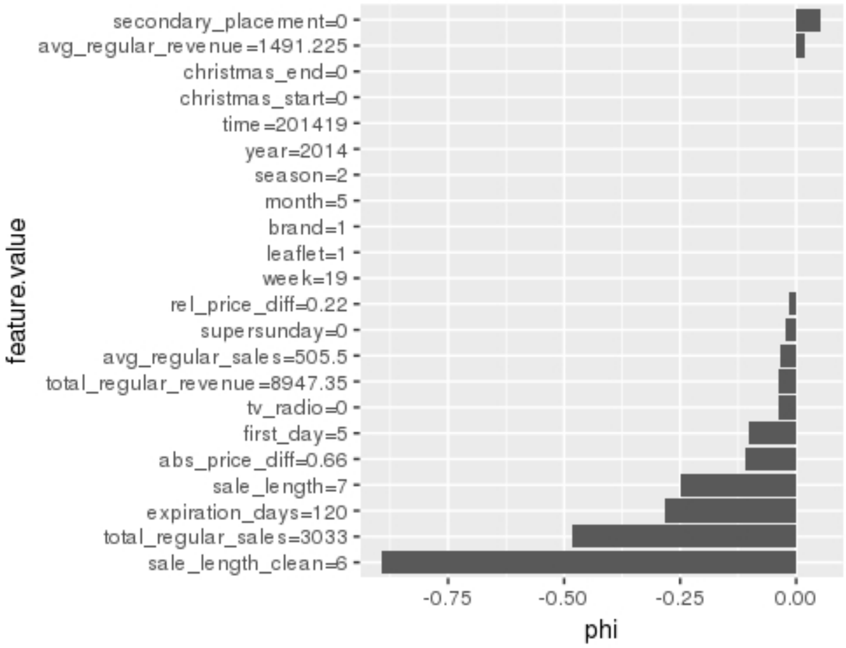


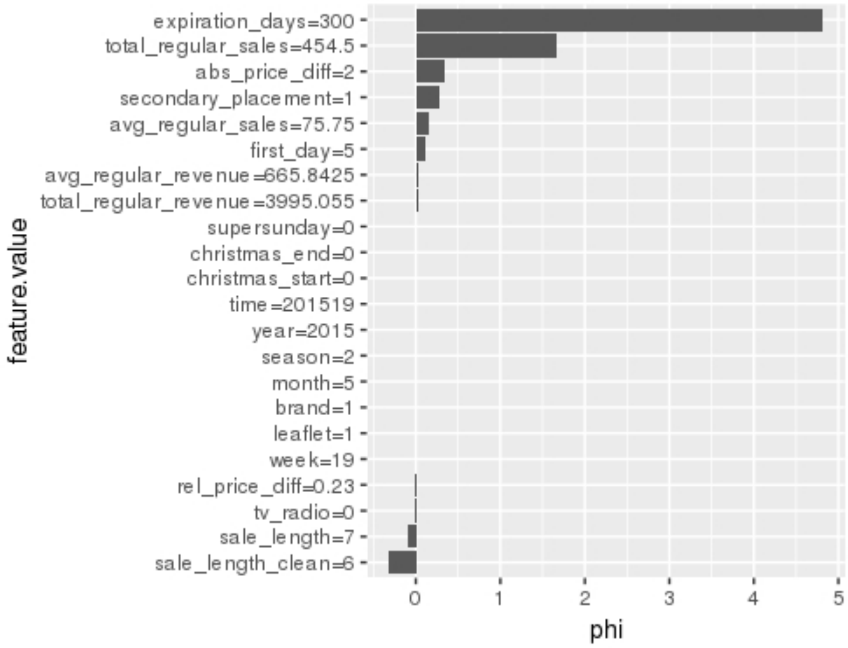
Regression LIME.



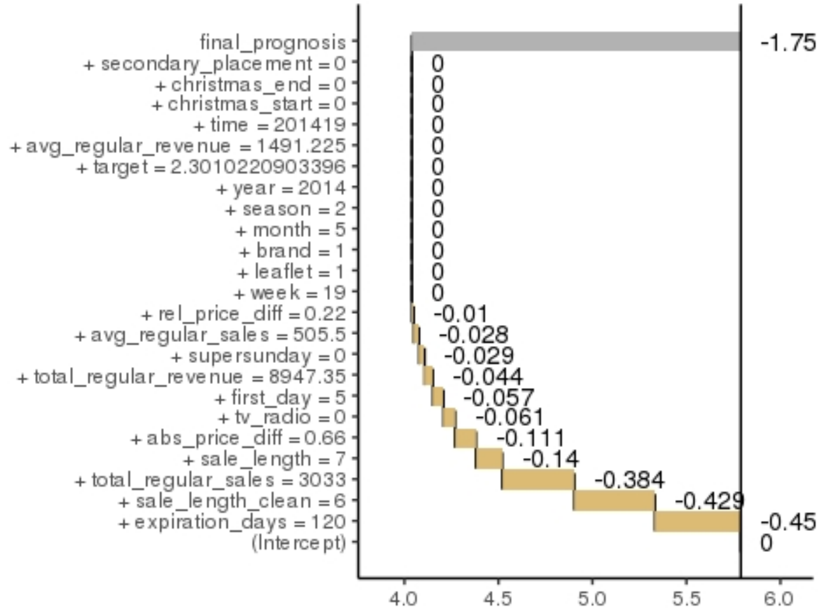


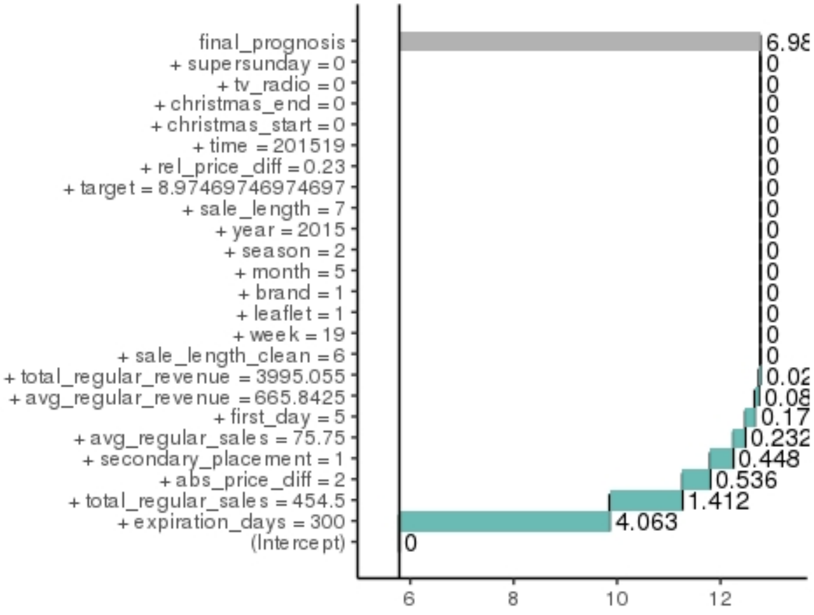
Regression. SHAP.



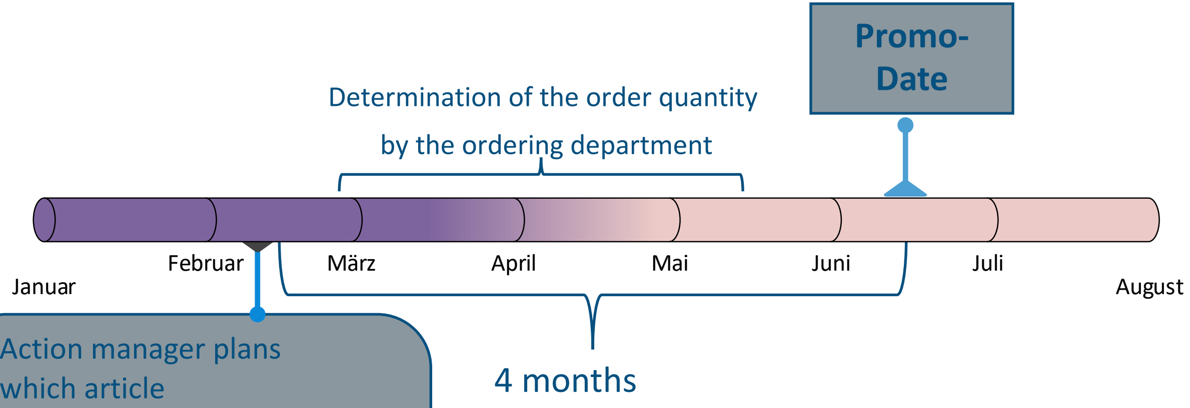


Regression. BreakDown.





How does it work?



- Action manager plans
  which article
  what discount
  which marketing measure
  with which duration
  is advertised

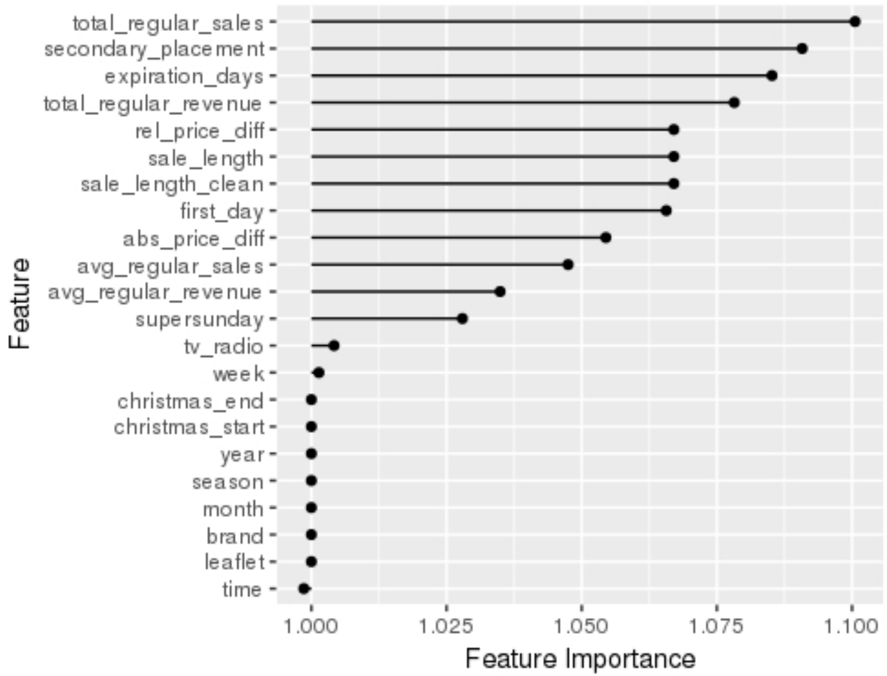
So far, the AM has been oriented towards the planning based on historical PVs.

Challenges for the action manager:

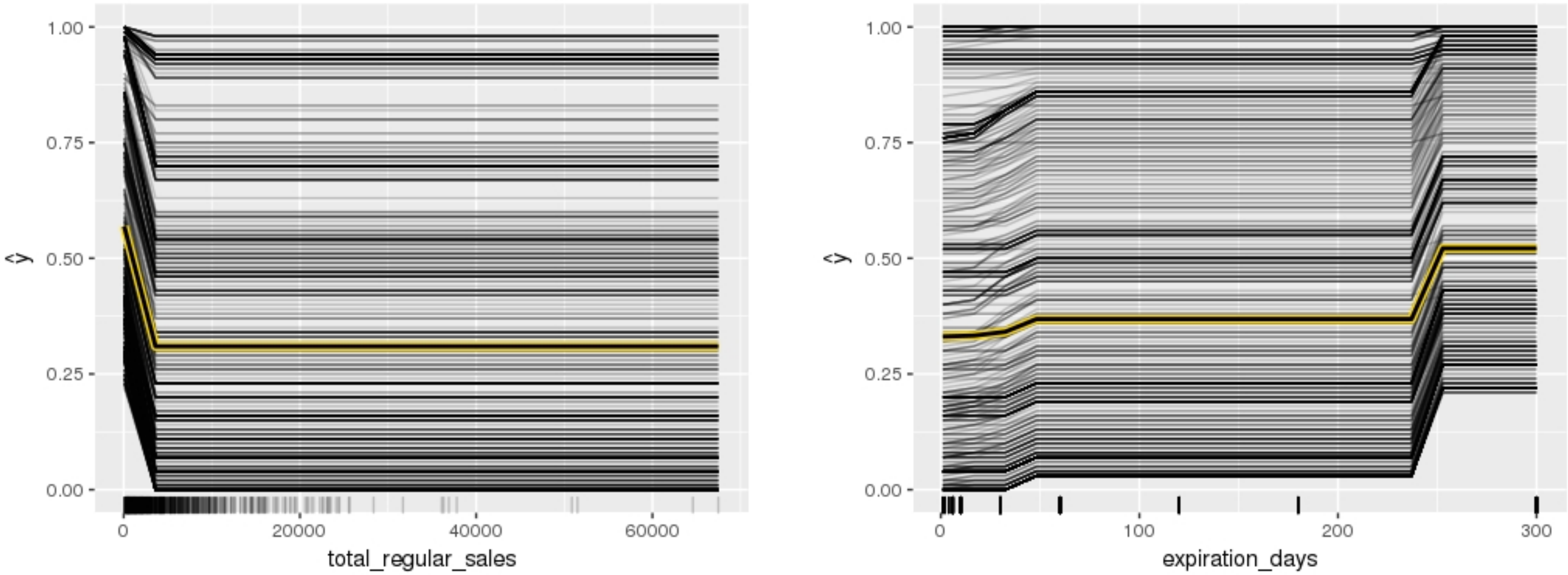
- Find the "best" comparable action manually
- Difficult with new articles / discount levels / duration / marketing measures etc.

Let‘s classify uplift into high vs low

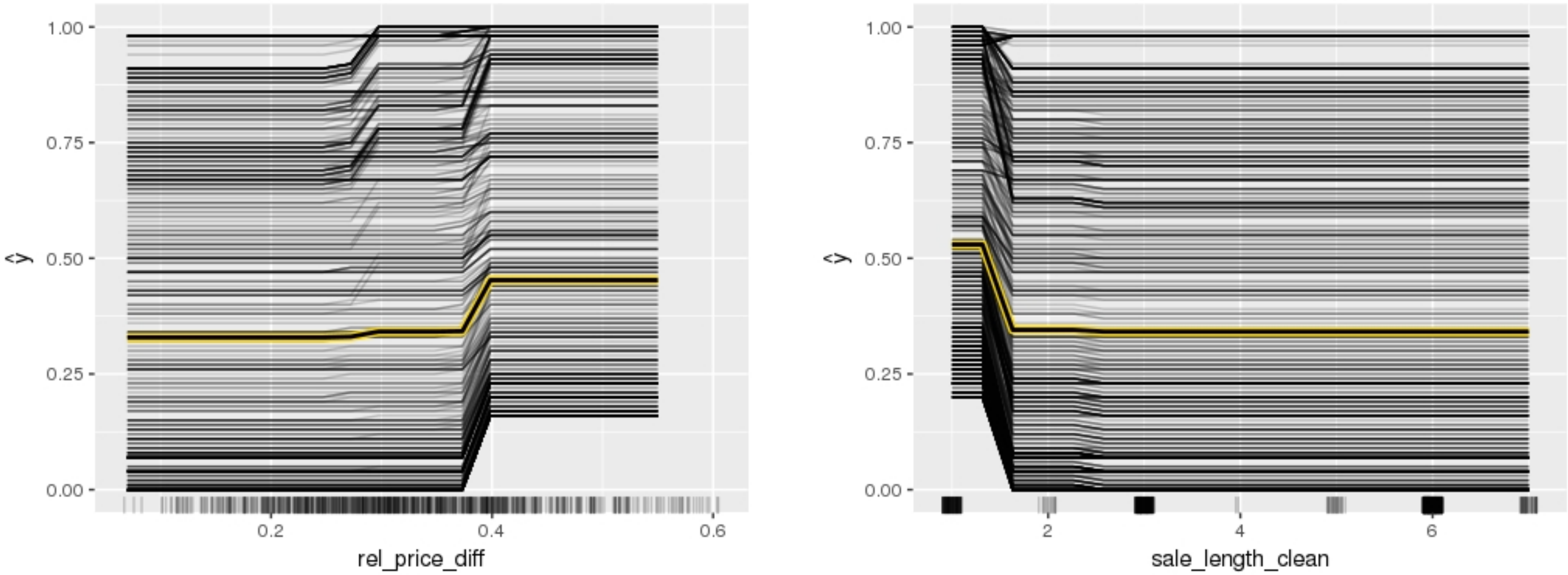
Classification. Feature Importance



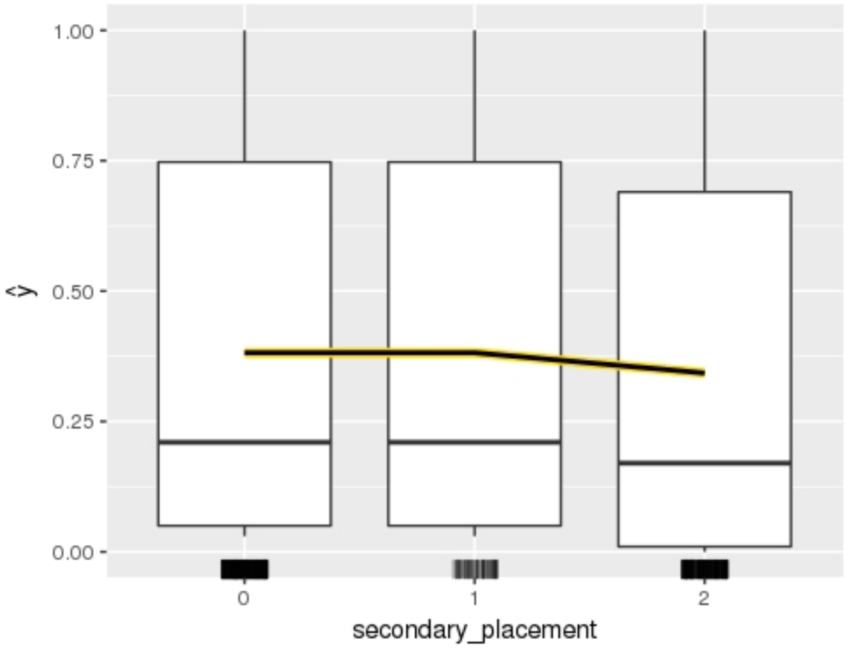
Classification. PDP.

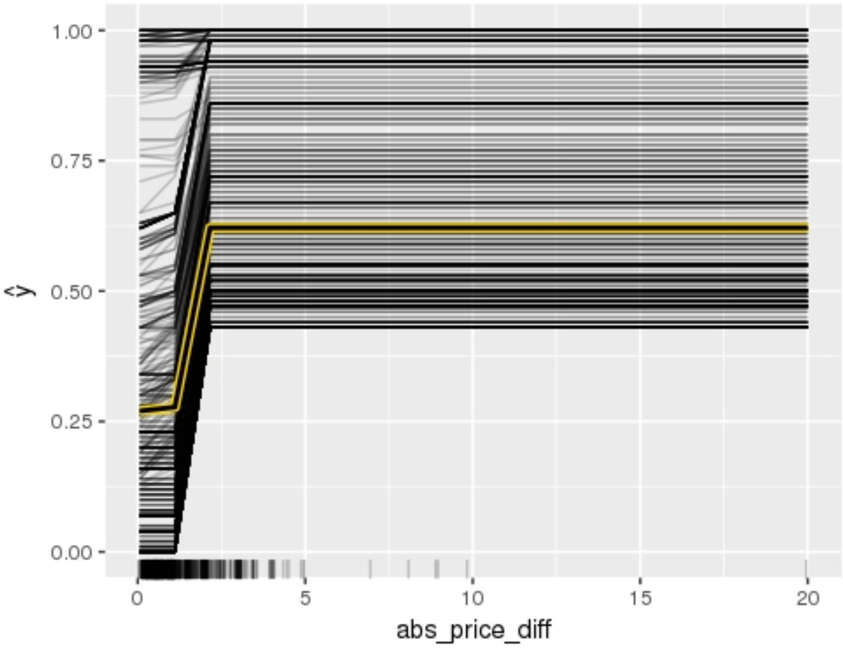


Classification. PDP.

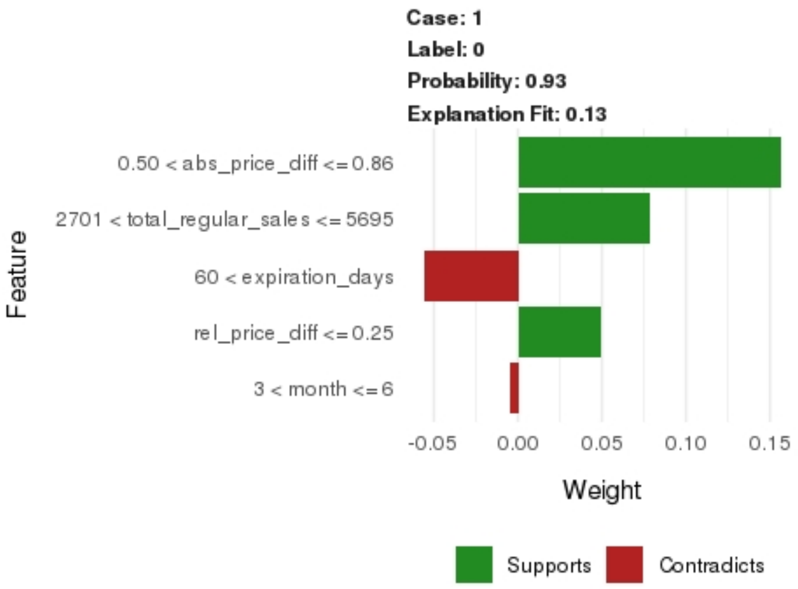


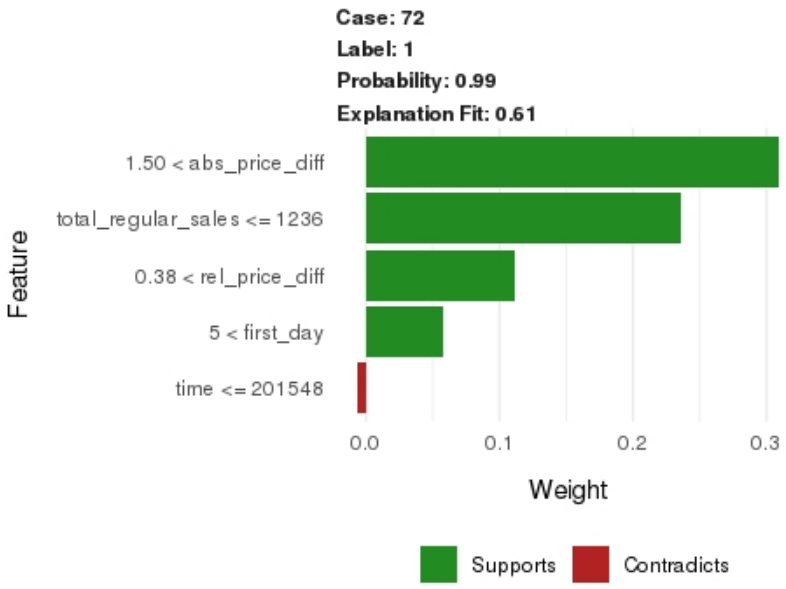
Classification. PDP.



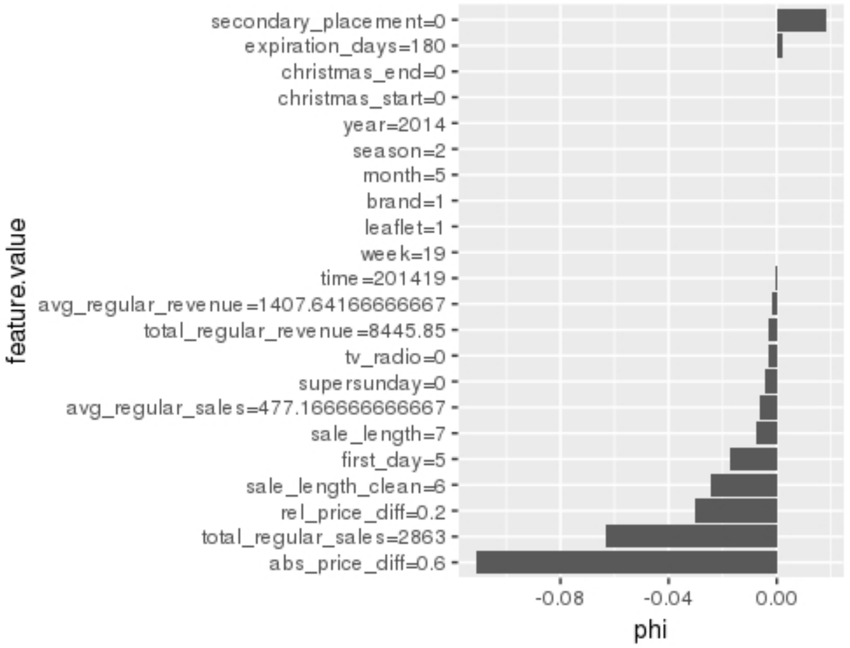


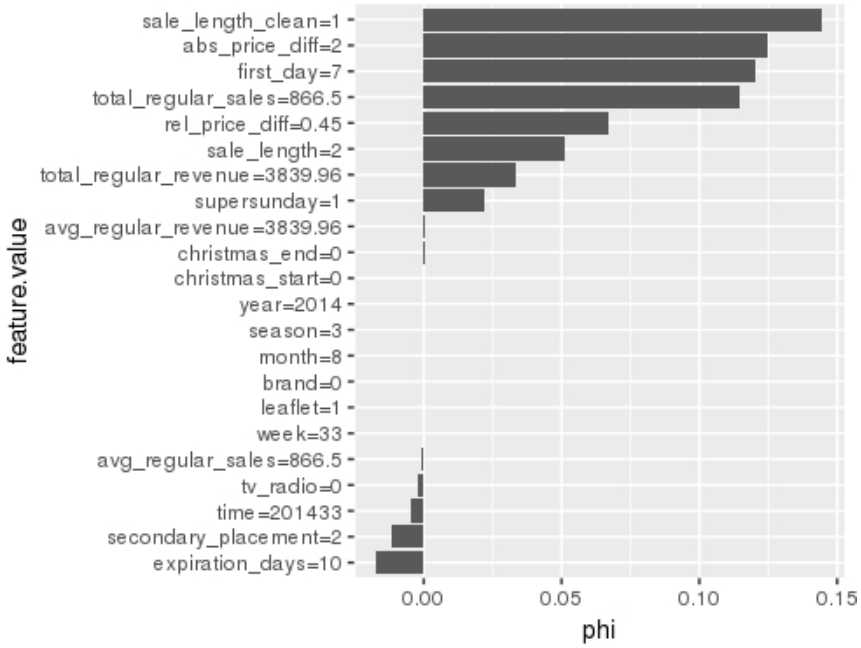
Classification. LIME.



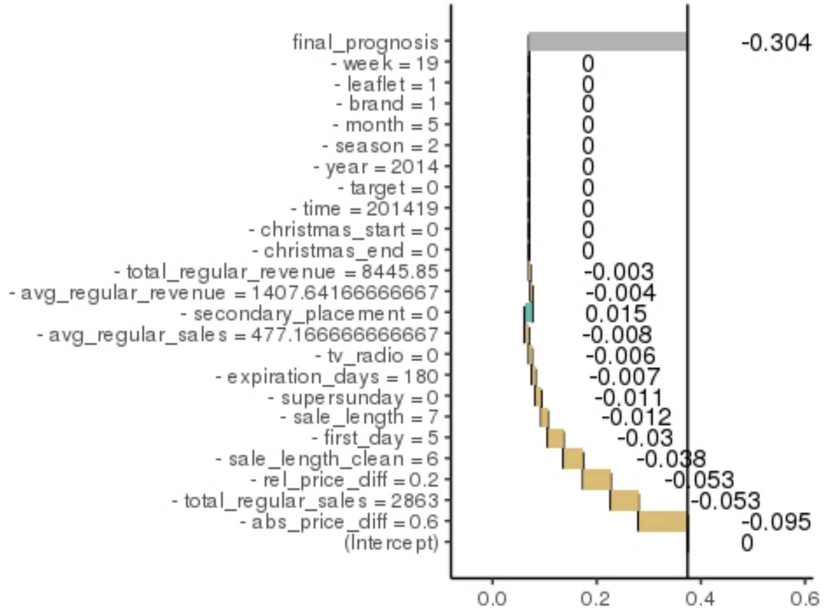


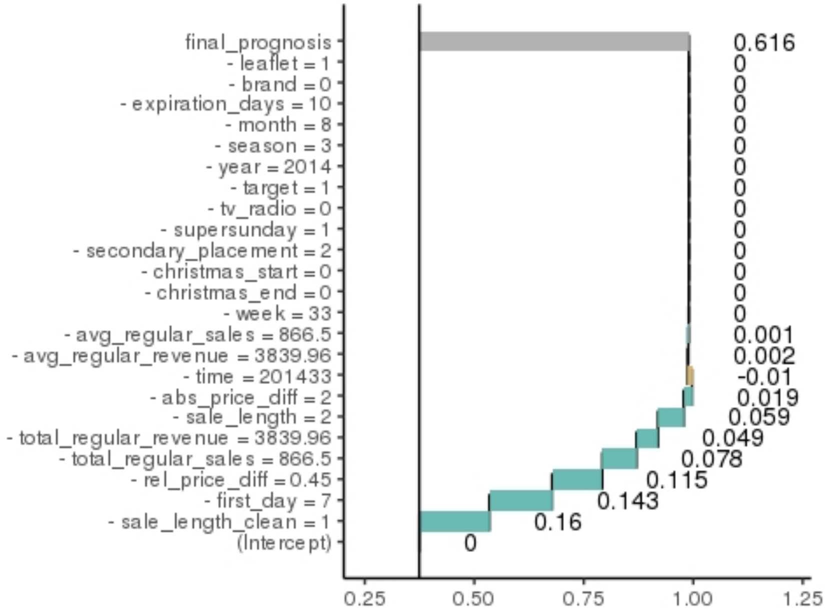
Classification. SHAP.





Classification. breakDown.





Can we explain each decision? Anchor.

```python
from __future__ import print_function
import numpy as np
import sys
import sklearn
import sklearn.ensemble
from anchor import utils
import utilsext
from anchor import anchor_tabular

dataset_folder = '/home/sharapov/anchor/datasets/'
dataset = utilsext.load_dataset_ext('pv', balance=True, dataset_folder=dataset_folder, discretize=True)

explainer = anchor_tabular.AnchorTabularExplainer(dataset.class_names, dataset.feature_names, dataset.data, dataset.categorical_names)
explainer.fit(dataset.train, dataset.labels_train, dataset.validation, dataset.labels_validation)

c = sklearn.ensemble.RandomForestClassifier(n_estimators=50, n_jobs=5)
c.fit(explainer.encoder.transform(dataset.train), dataset.labels_train)
predict_fn = lambda x: c.predict(explainer.encoder.transform(x))
print('Train', sklearn.metrics.accuracy_score(dataset.labels_train, predict_fn(dataset.train)))
print('Test', sklearn.metrics.accuracy_score(dataset.labels_test, predict_fn(dataset.test)))

idx = 1
np.random.seed(1)
print('Prediction: ', explainer.class_names[predict_fn(dataset.test[idx].reshape(1, -1))[0]])
exp = explainer.explain_instance(dataset.test[idx], c.predict, threshold=0.95)

print('Anchor: %s' % (' AND '.join(exp.names())))
print('Precision: %.2f' % exp.precision())
print('Coverage: %.2f' % exp.coverage())
```

Source https://github.com/marcotcr/anchor-experiments

Anchor.

- **Anchor:** 1.80 < **abs_price_diff** < 3.11 **AND** 327.75 < **avg_regular_sales** < 662.83
- **Anchor: expiration_days** = 6 **AND sale_length_clean** = 3 **AND** 4689.8 < **total_regular_revenue** < 11219.6

Contrastive explanation

- The model predicted 5.48 instead of more than 5.48 because **abs_price_diff** <= 1.796 **and year** > 2015
- The model predicted 4.54 instead of more than 4.54 because **total_regular_revenue** > 6129.841 **and sale_length** > 6.463
- The model predicted 5.17 instead of more than 5.17 because **not christmas_end and rel_price_diff** <= 0.369

Looking forward.

- XAI will gain more and more popularity in industry since
  - stakeholders and business leader demand explainability of ML models and results
  - doctors, lawyers will be use it to justify their own decisions
  - Fair ML and XAI will go hand in hand.
  - applications to image and test data

Future of XAI

- Accurate Models
- Trustworthy Models
- Natural Language Explanation
- Adversarial Use (misuse)
- Collaboration with Machine

Join the club

- https://www.linkedin.com/groups/8672810/ - my LinkedIn group where I post XAI related article.
- You can find here a lot of useful links and community interested in XAI.
- You can also contact me directly on LinkedIn.

Software

Python

- aequitas
- anchor
- ContrastiveExplanation (Foil Trees)
- eli5
- fairml
- L2X
- lime
- pyBreakDown
- PDPbox
- PyCEbox
- shap
- Skater
- tensorflow/model-analysis
- themis-ml
- treeinterpreter

R

- ALEPlot
- breakDown
- DALEX
- ExplainPrediction
- ICEbox
- iml
- lime
- live
- xgboostExplainer
- lightgbmExplainer

What do we offer.

- Positions for interships required by university or for bachelor/master thesis

- Voluntary internships in general (including compensation)

- Working student contracts

Contact.

andrey.sharapov@lidl.com

Thank you