Black is the new White

using eXplainable Artificial Intelligence in Business

University of Warsaw
Faculty of Economic Sciences

DATA DONUTS

Data Juice Lab.

Biznes oparty na danych

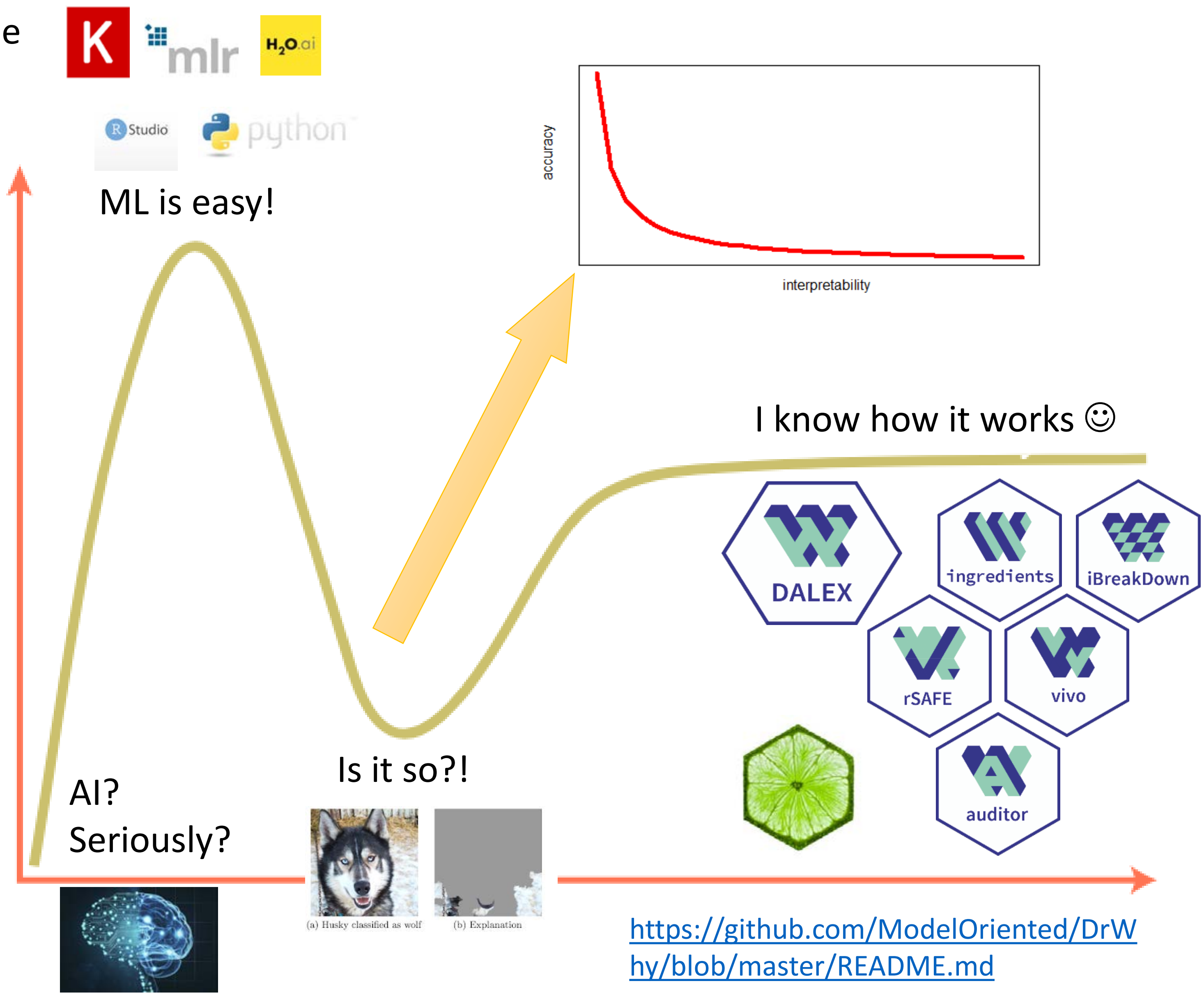
data

Agenda

1. DO I NEED XAI?
2. DO I NEED BLACK-BOXES?
3. HOW CAN I USE XAI IN BUSINESS?

1. DO I NEED XAI? (1/4)

Hype cycle

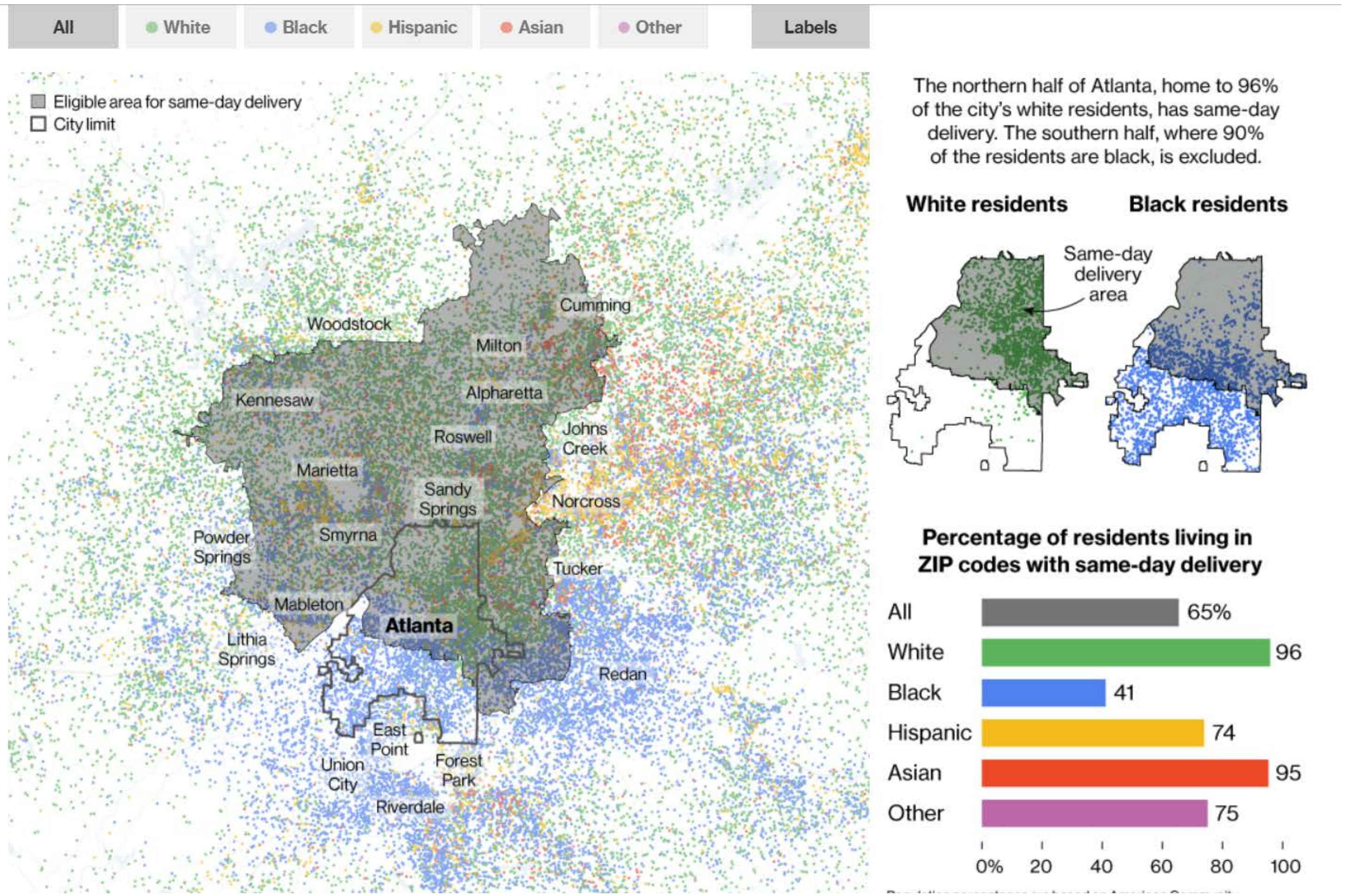


https://github.com/ModelOriented/DrWhy/blob/master/README.md

1. DO I NEED XAI? (2/4)

ML failures (1/2)

- In 2016 Amazon bulit a ML model that suport same-day-deivery process
- Unintentaionally model discirminate etnhnic gorups



https://www.bloomberg.com/graphics/2016-amazon-same-day/

1. DO I NEED XAI? (3/4)

ML failures (2/2)

THE VERGE TECH SCIENCE CULTURE CARS REVIEWS LONGFORM VIDEO MORE

TECH AMAZON ARTIFICIAL INTELLIGENCE

21

Amazon reportedly scraps internal AI recruiting tool that was biased against women

The secret program penalized applications that contained the word "women's"

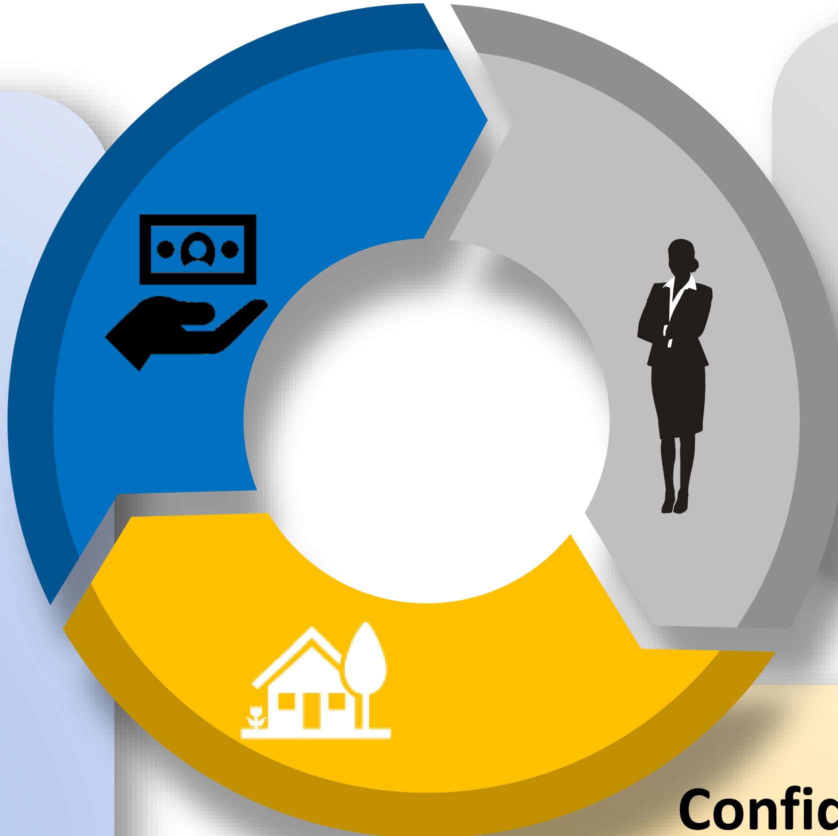
By James Vincent | @jjvincent | Oct 10, 2018, 7:09am EDT

In effect, Amazon's system taught itself that male candidates were preferable. It penalized resumes that included the word "women's," as in "women's chess club captain." And it downgraded graduates of two all-women's colleges, according to people familiar with the matter. They did not specify the names of the schools.

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

1. DO I NEED XAI? (4/4)

Can we trust the model?



Business

- Reason for a business decision
- Importance of a features
- What-if analyses
- Avoiding non-ethical decisions
- Fraud detection

Regulations

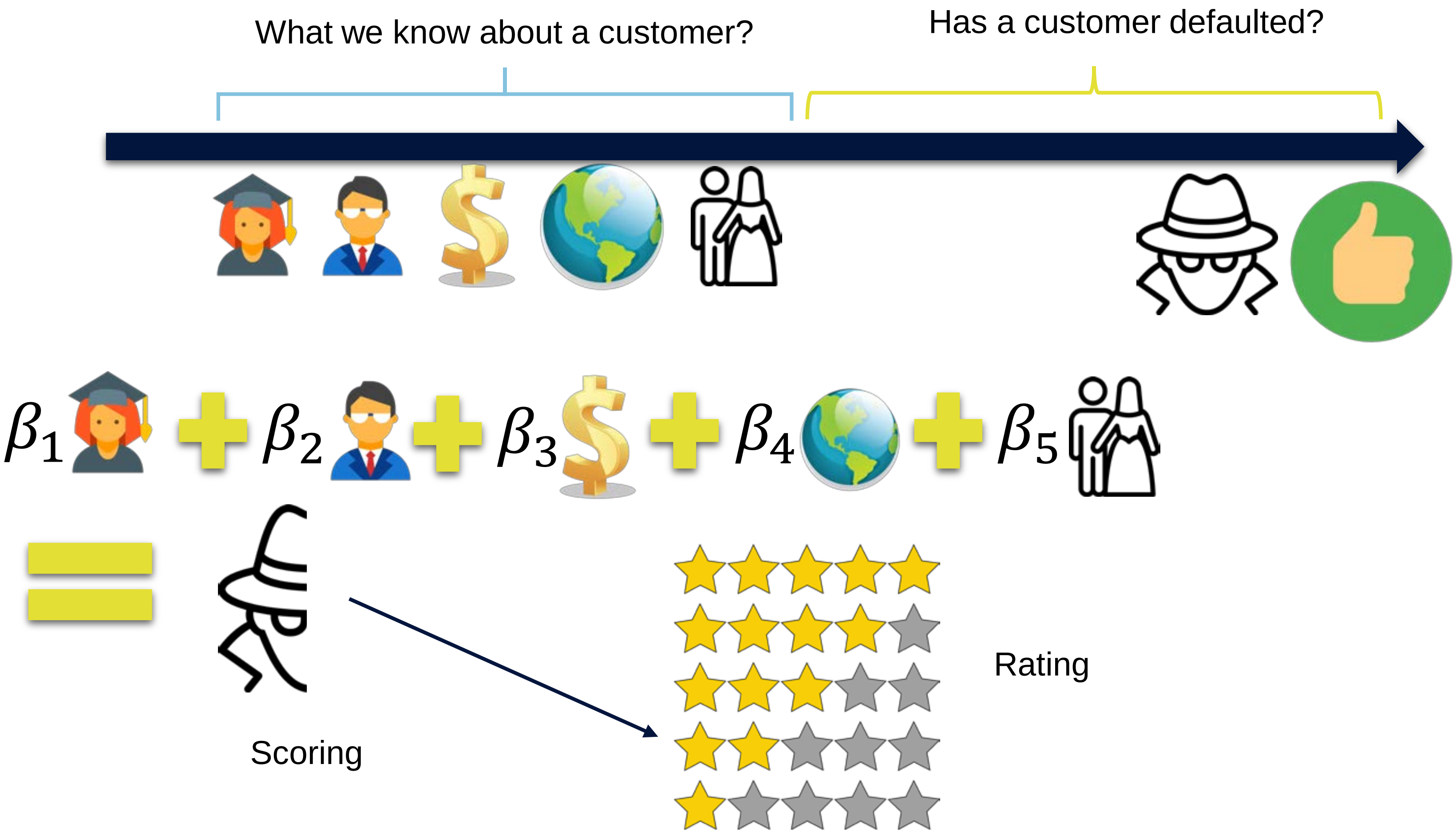
- GPDR - European Parliament (2018)
  - Trustworthy AI;
  - ethics, discrimination
- Right to know (2019) GPDR
- Know your model (`00)

Confidence

- Intuition of results
- Stability of results
- Confidence of results

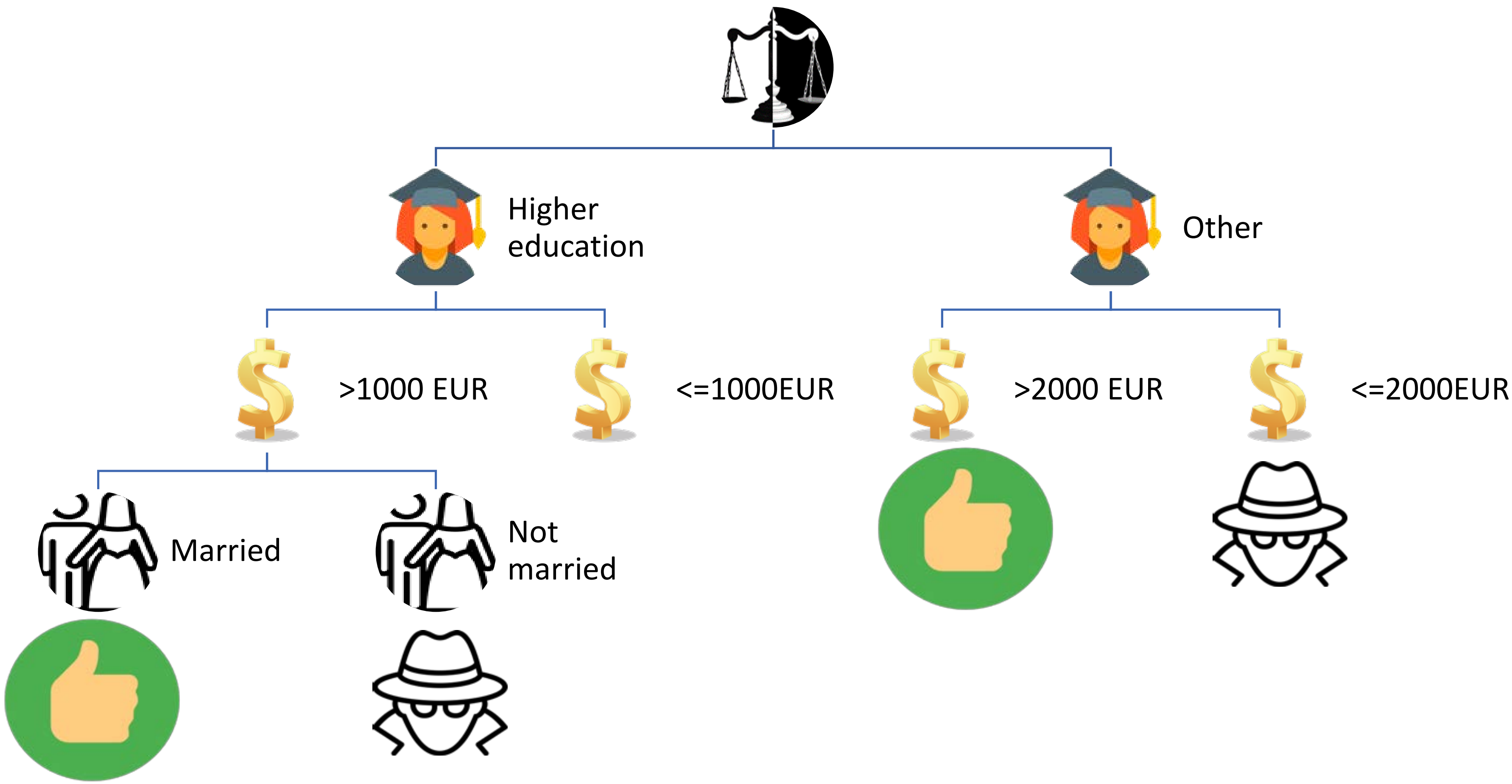
2. DO I NEED BLACK-BOXES? (1/7)

Logistic Regression



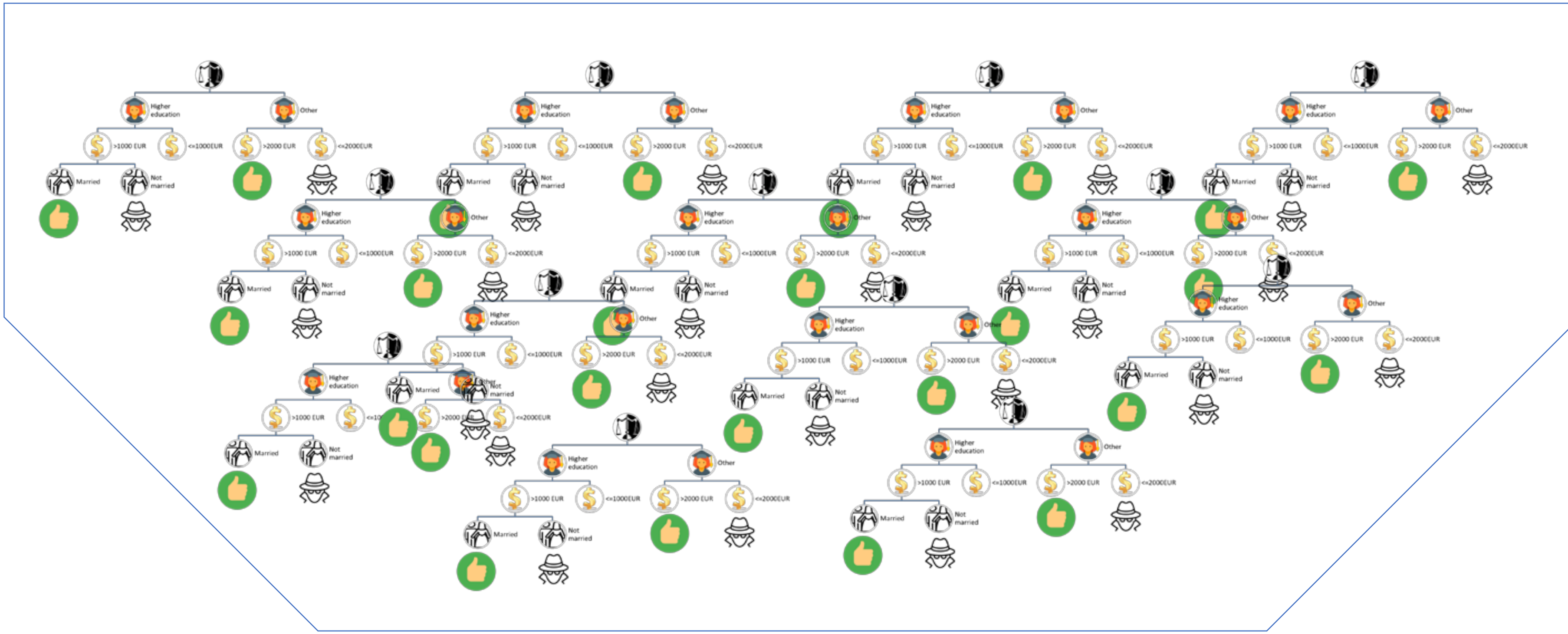
2. DO I NEED BLACK-BOXES? (2/7)

Decision Tree



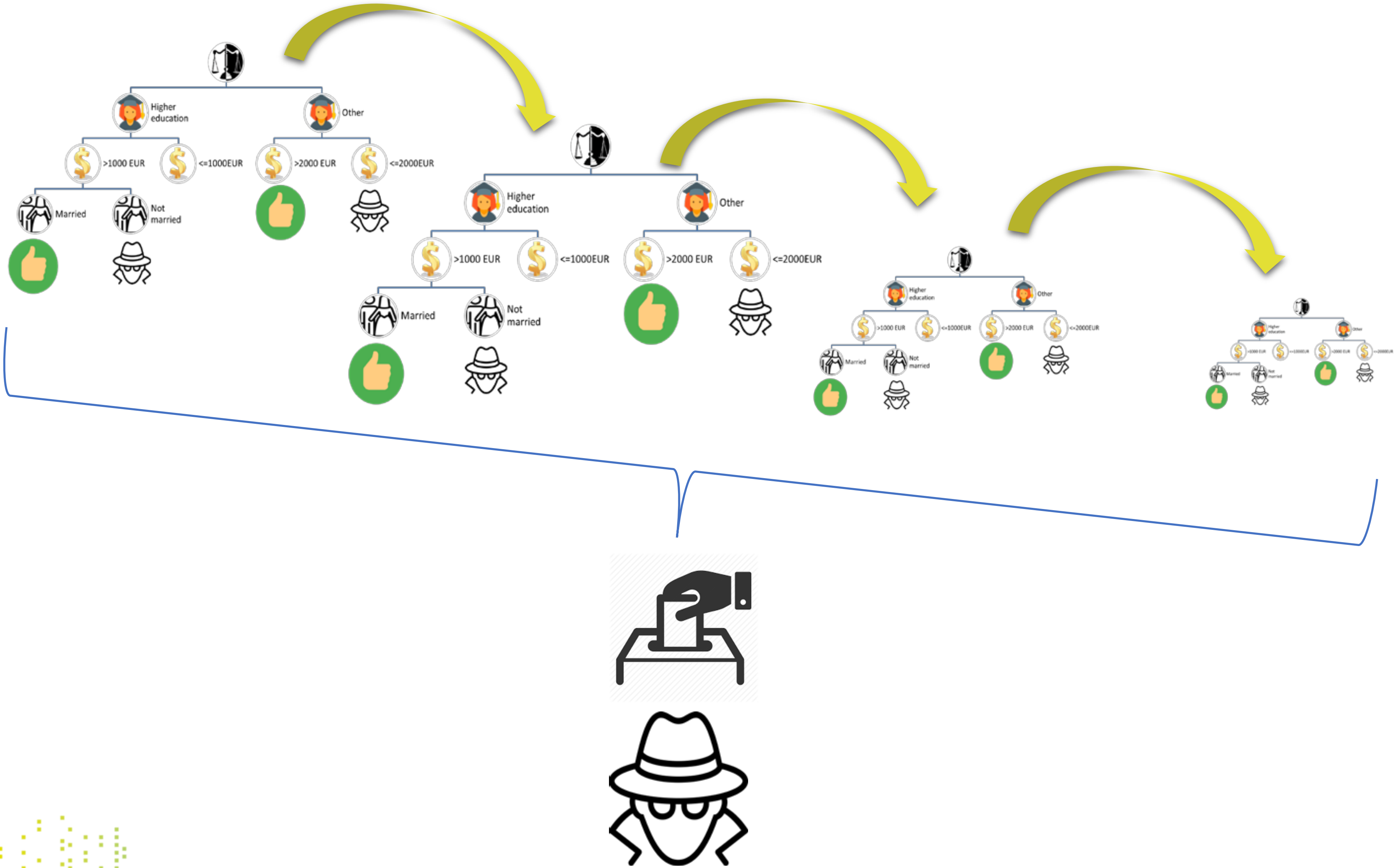
2. DO I NEED BLACK-BOXES? (3/7)

Random Forest



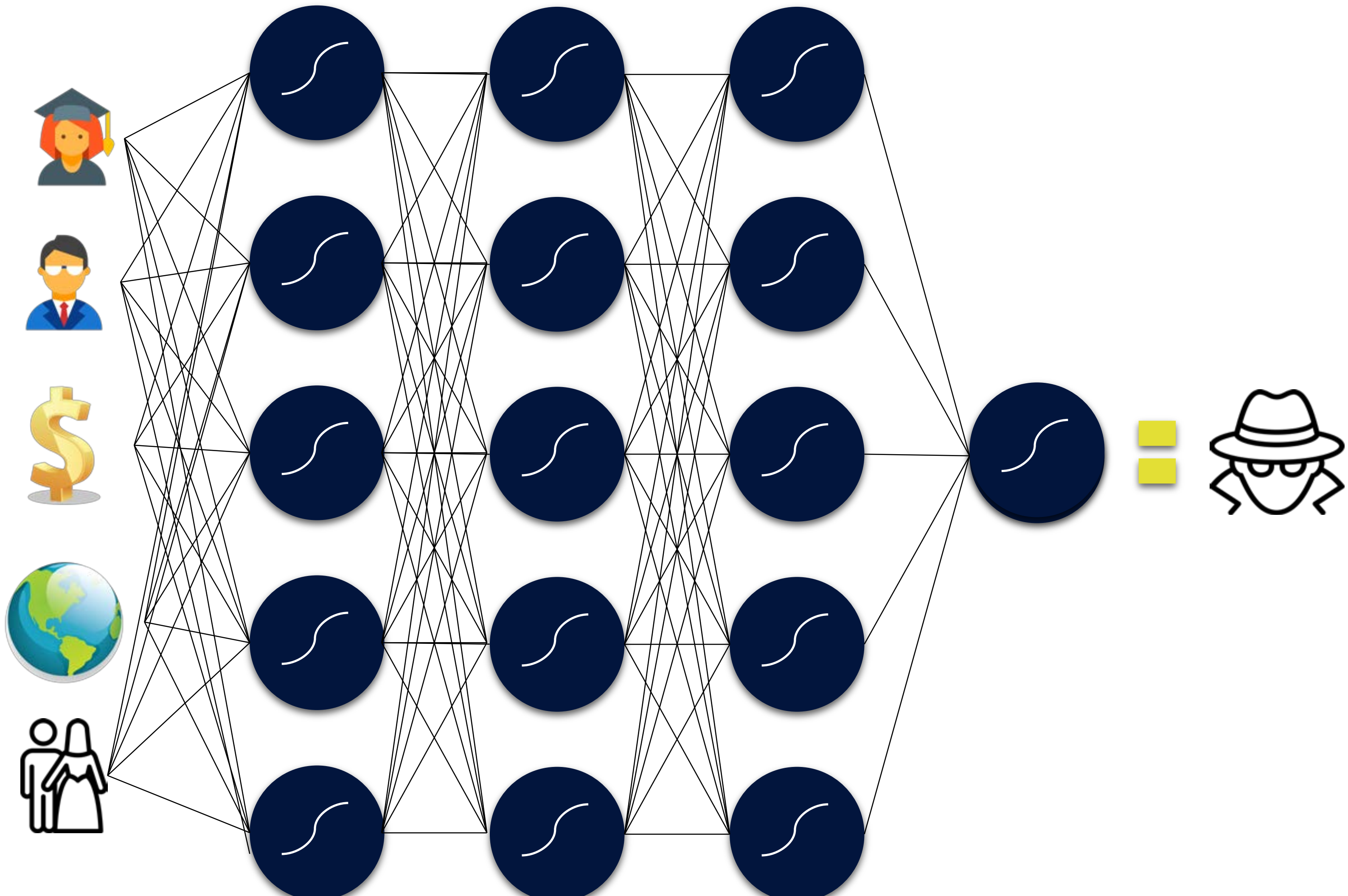
2. DO I NEED BLACK-BOXES? (4/7)

Boosted trees



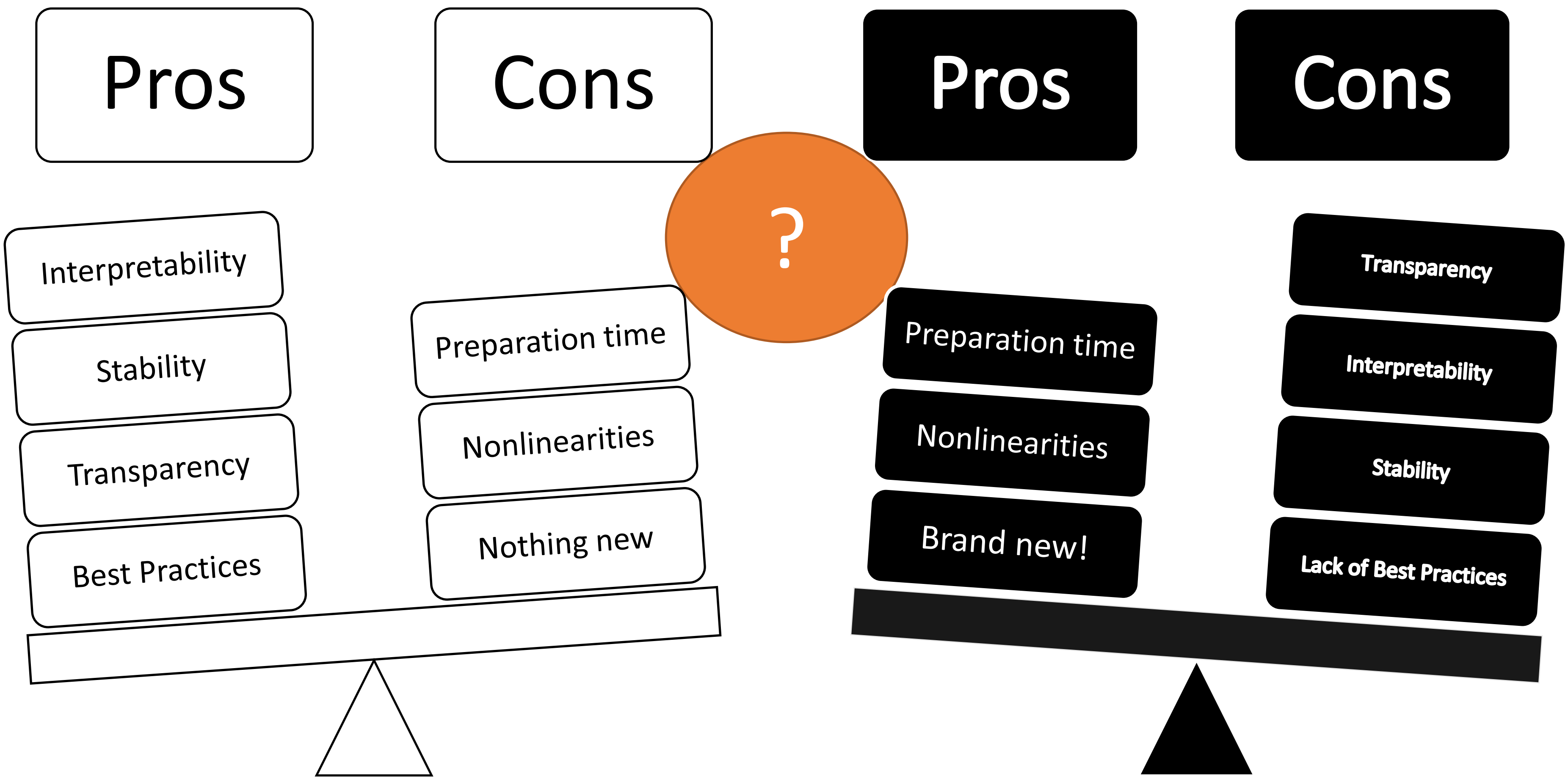
2. DO I NEED BLACK-BOXES? (5/7)

Neural Network

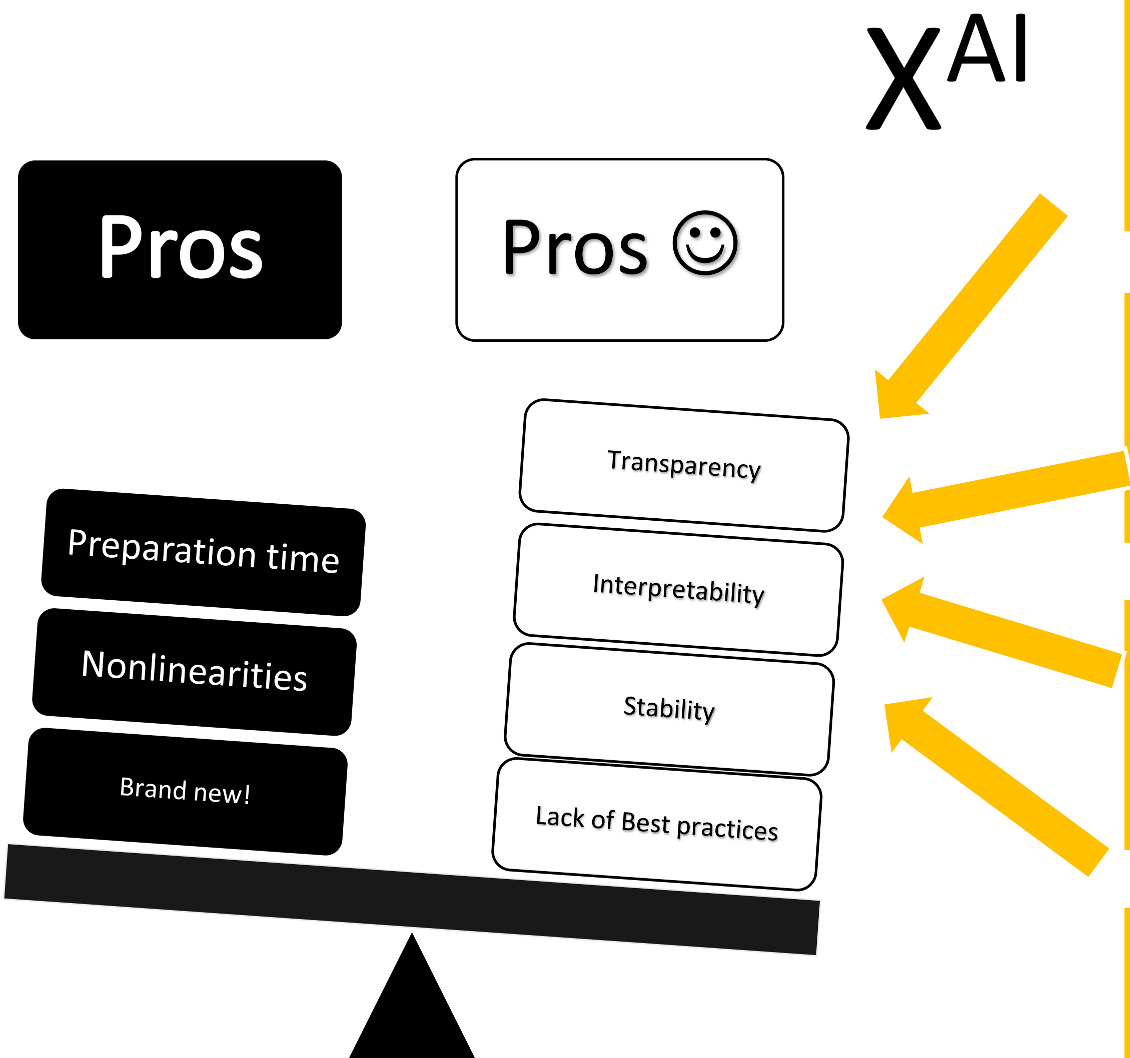


2. DO I NEED BLACK-BOXES? (6/7)

White (Glass) Box vs Black Box



2. DO I NEED BLACK-BOXES? (7/7)



How model work is becoming transparent

Local and Global explanation

Local and Global stability analysis

Agnostic approach
(new Best Practices)

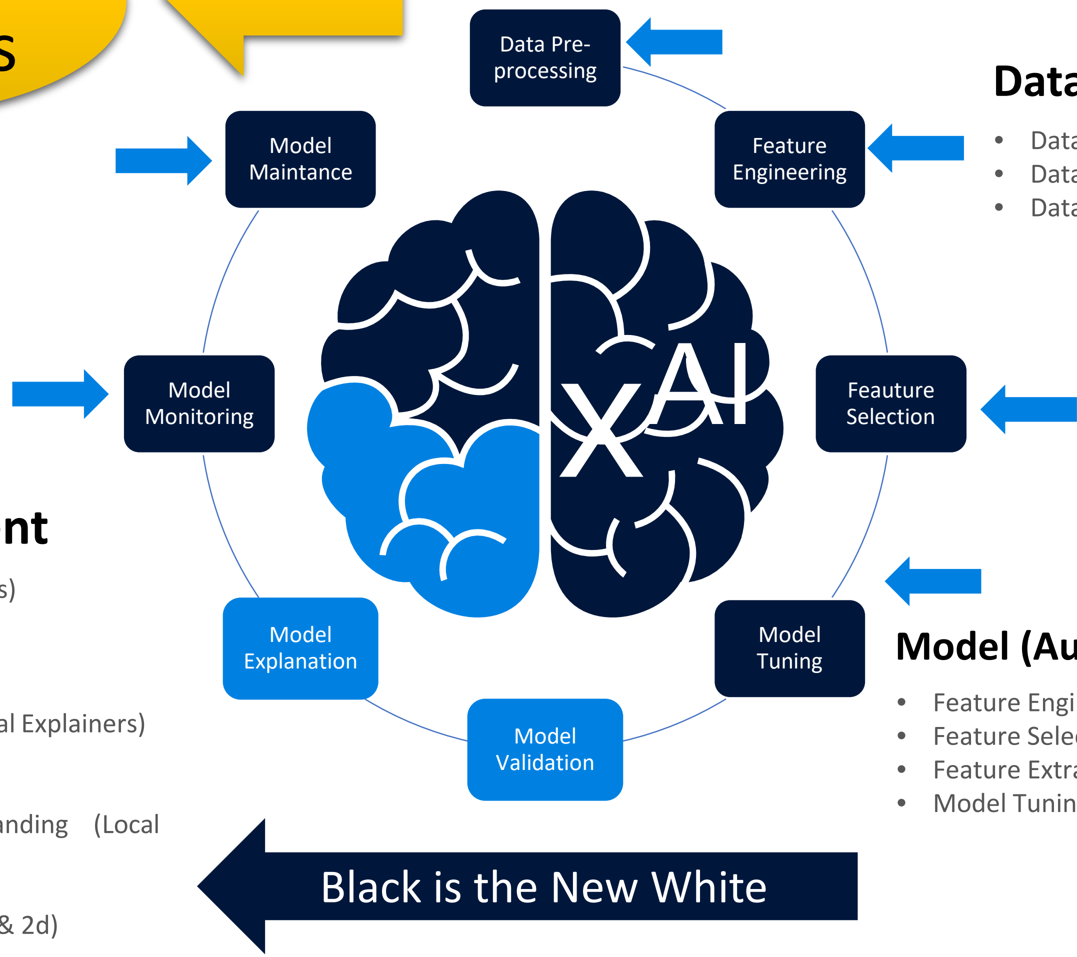
University of Warsaw
Faculty of Economic Sciences

DATA DONUTS

Data Juice Lab.
Biznes oparty na danych

3. HOW CAN I USE XAI IN BUSINESS? (1/16)

Business Decisions



Model Usage

- Model Monitoring
- Model Maintenance

Data pre-processing

- Data acquisition
- Data cleansing
- Data pre-analysis

Model Assessment

- Model Performance (Auditors)
  - Discriminatory
  - Stability
  - Error analysis
- Variables Performance (Global Explainers)
  - Joint analysis
  - Case-by-case
- Single prediction understanding (Local Explainers)
  - Variables role
  - What-if analysis (1d & 2d)

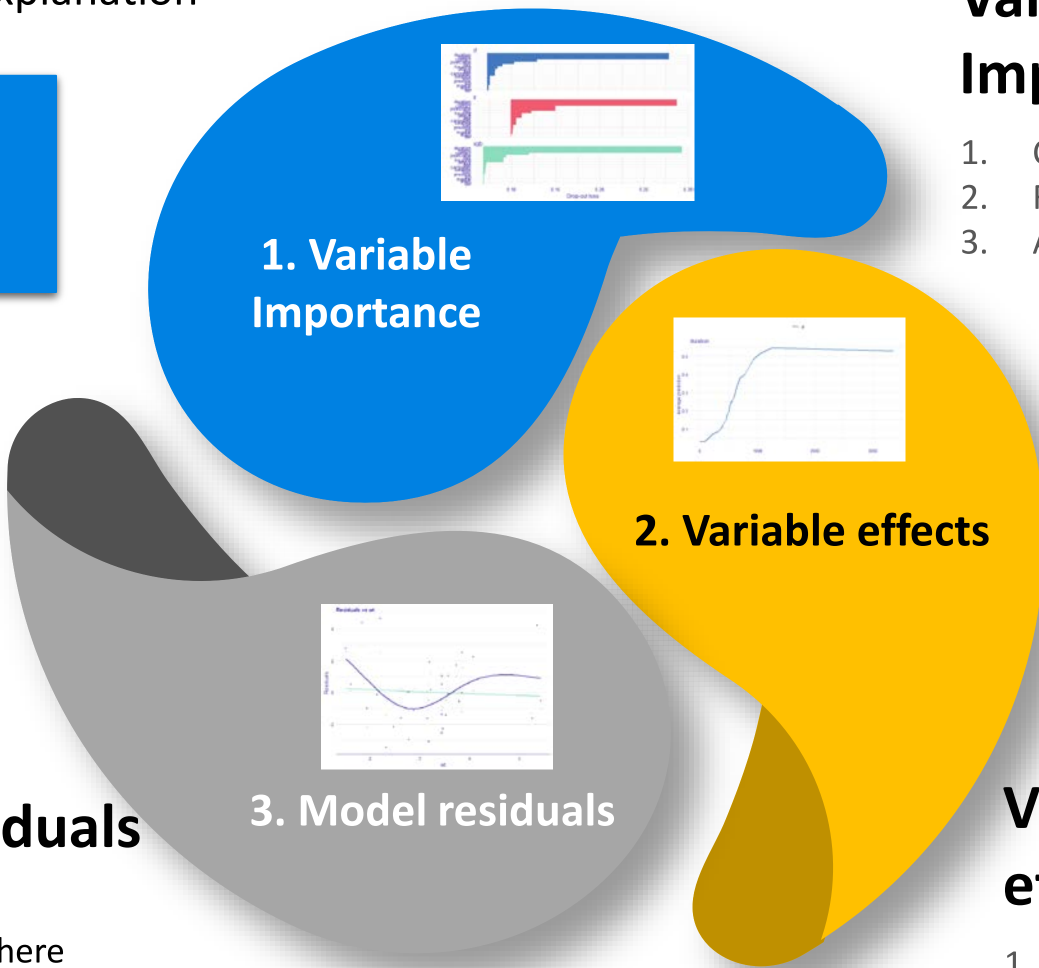
Model (Auto)Development

- Feature Engineering
- Feature Selection
- Feature Extraction
- Model Tuning

3. HOW CAN I USE XAI IN BUSINESS? (2/16)

Global and Loca Explanation

Model performance



Variable Importance

1. Confirmation wtr intuition
2. Reduction of complexity
3. Aspects identification

Model residuals

1. Identification of subpopulations where model is performing poorly
2. Non-random residuals patterns

Variables effects

1. Confirmation wrt intuition
2. Feature engineering
3. Getting knowledge about phenomenon
4. Model selection

3. HOW CAN I USE XAI IN BUSINESS? (3/16)

Dalex explainer

Black-box models may have very different structures. *explain()* creates a unified representation of a model, which can be further processed by various explainers.

Key parmeters:

- model – the model that we want to explain
- data – data that we want to use to explain the model
- y – target variable
- predict_function – function that based on model and dataset provide predictions
- residual_function - function that based on model, dataset and target provide residuals
- label – name of an explainer

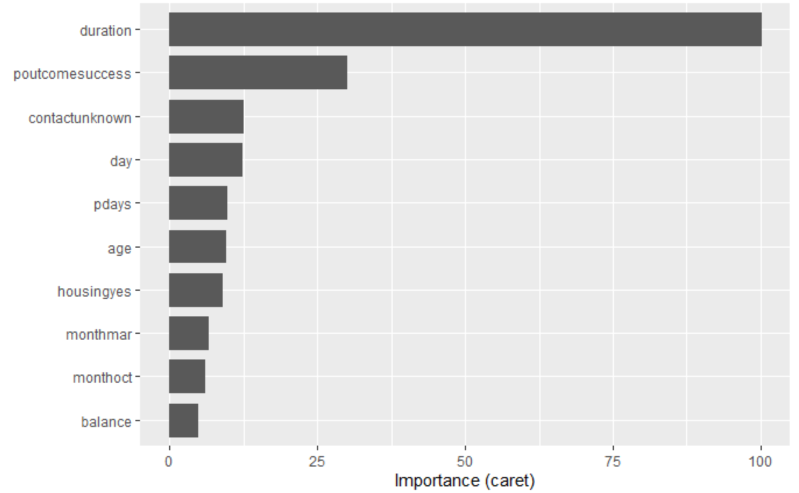
https://github.com/ModelOriented/DALEX

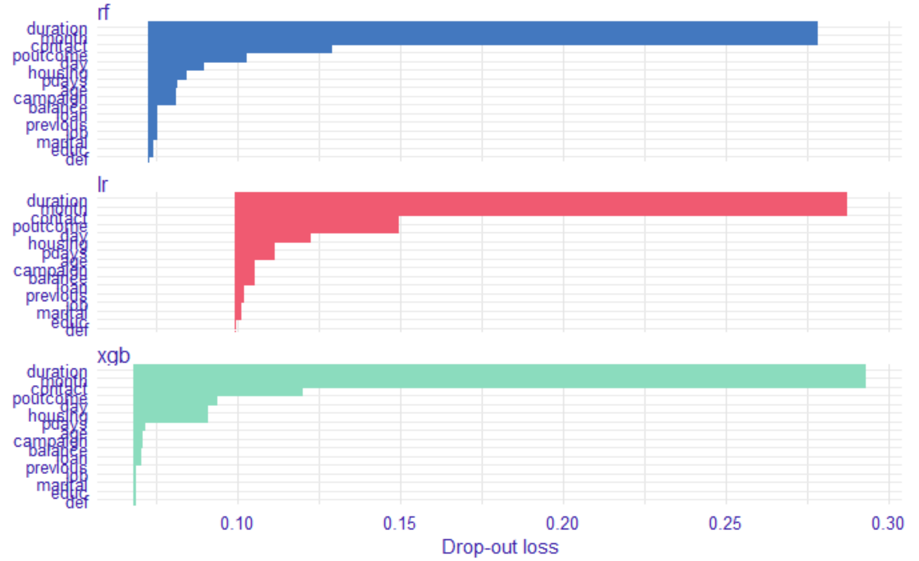
https://pbiecek.github.io/PM_VEE/

3. HOW CAN I USE XAI IN BUSINESS? (4/16)

Feature importance

- A group of methods that help to assess the role of the variables in the model.
- Model specific (Caret) and model agnostic methods (Ingredients)

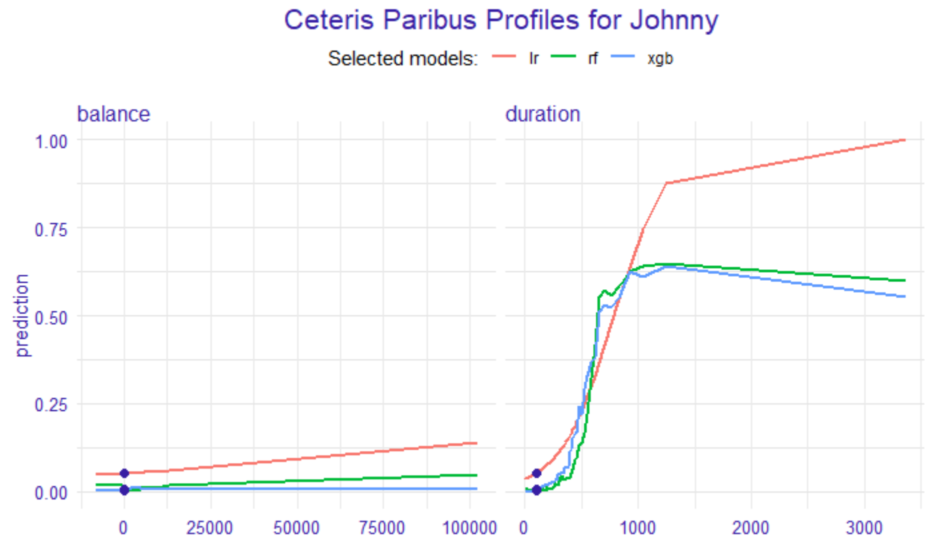




3. HOW CAN I USE XAI IN BUSINESS? (5/16)

Ceteris Paribus Plot

- In essence, a CP profile shows a conditional expectation of the dependent variable (response) for the particular explanatory variable.
- For a given observations all variables except one are kept, the one is changing and prediction of depvar is calculated



Pros:

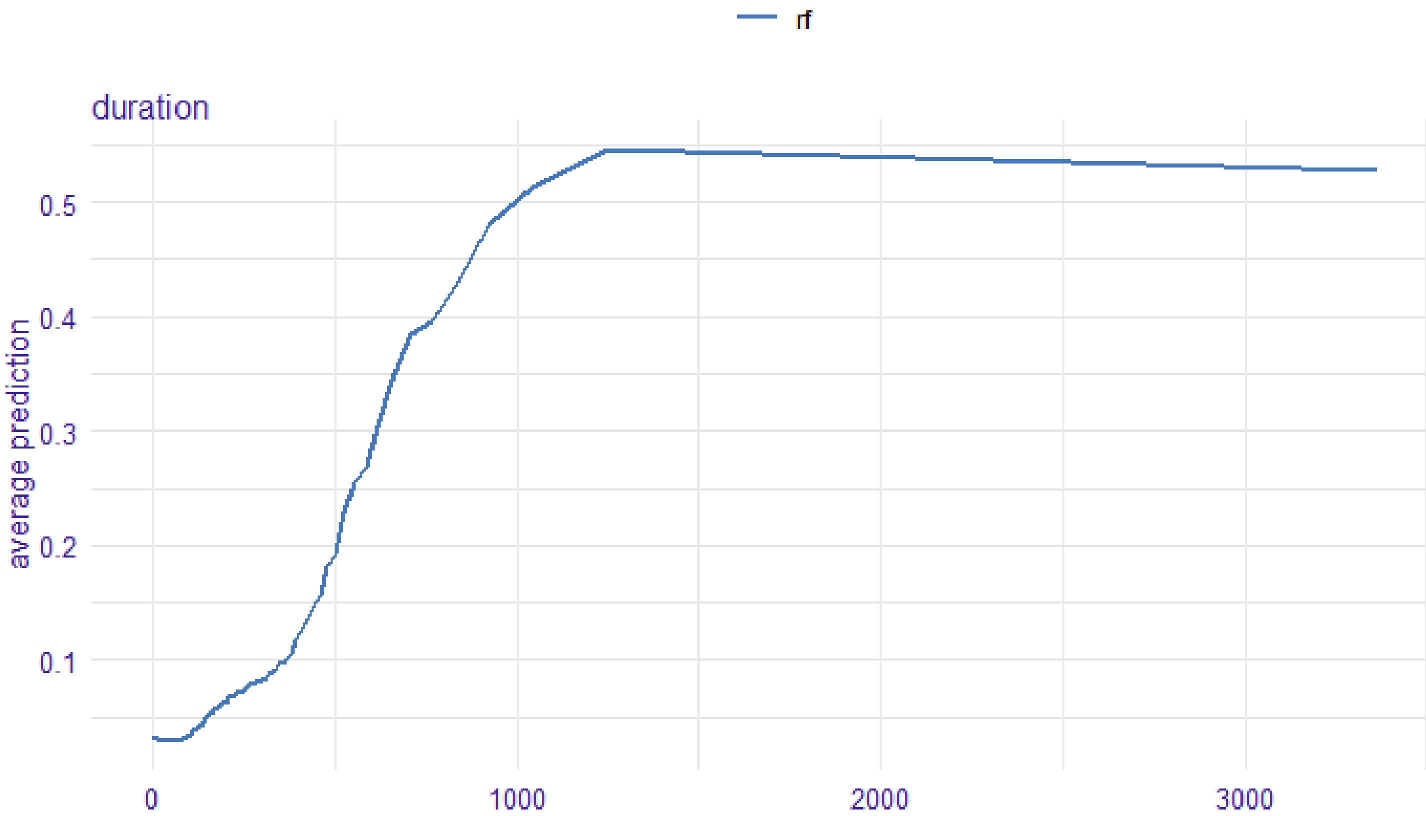
1. Uniform and easy to understand concept
2. Many profiles in a single plot.
3. Easy to compare

Cons:

1. Collinearity of indepvars
2. Factors with many levels

Partial Dependence Profiles

Feature effect analysis – average over CP for different instances



Pros:

- easy to interpret and understand.

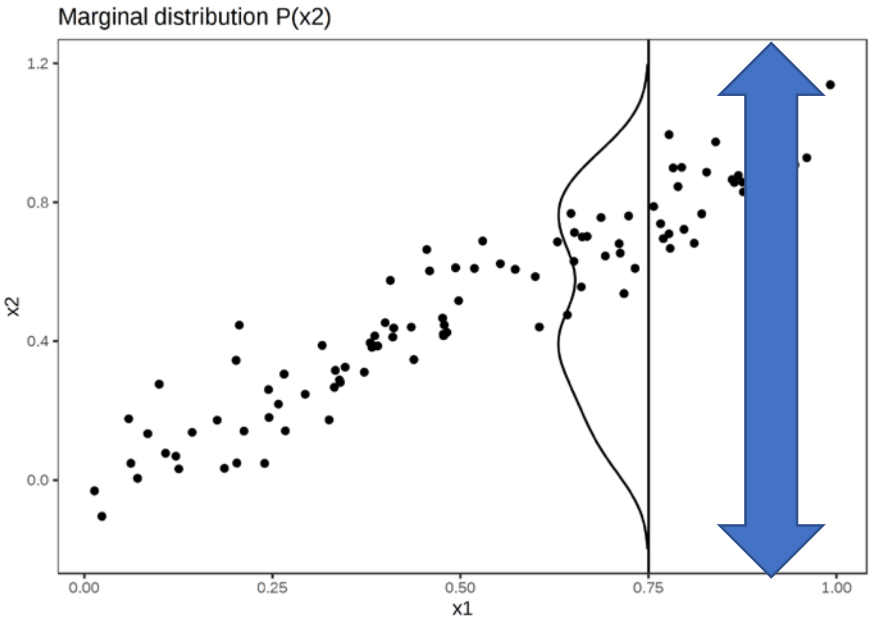
Cons:

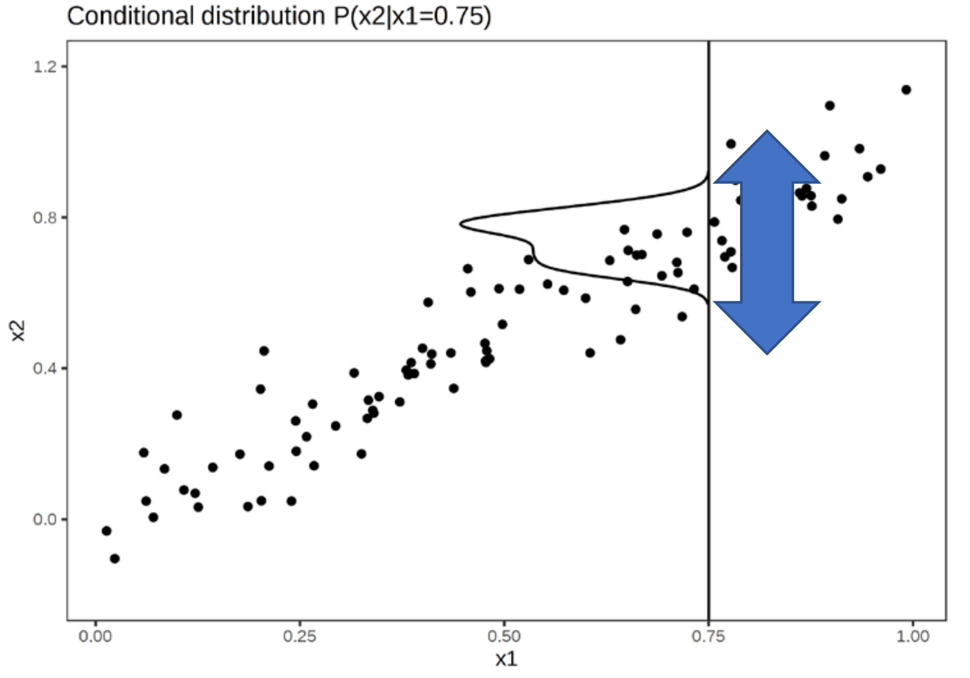
- Independency between indepvars
- Calculation over non-existed observations

3. HOW CAN I USE XAI IN BUSINESS? (7/16)

Conditional Dependency Profiles

- Instead of marginal distribution over X2 a conditional distribution of X2 is used
- Pros: Include information about correlated indepvar,
- Cons: report spurious relations via correlation (correlated vars effect)

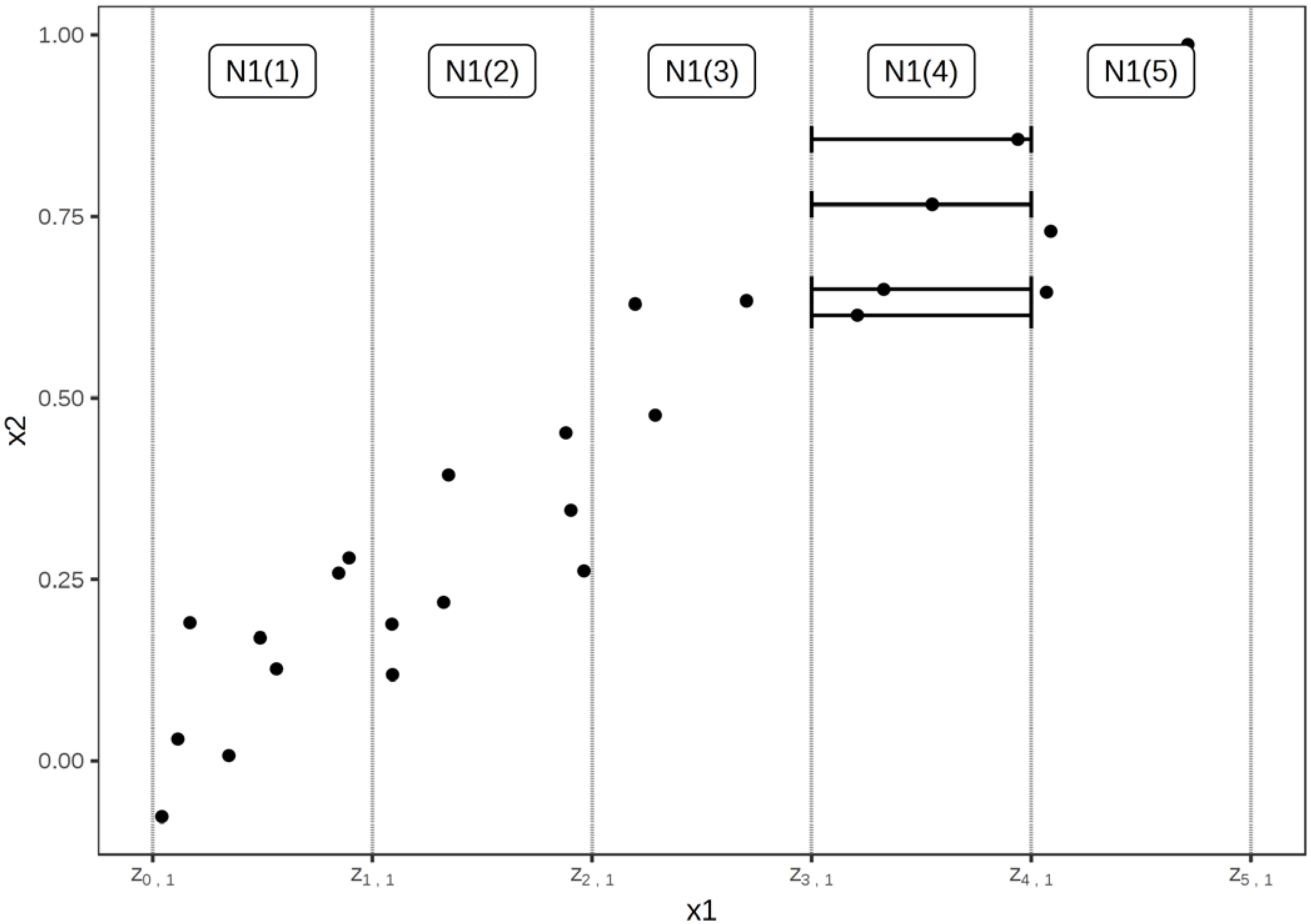




https://christophm.github.io/interpretable-ml-book/

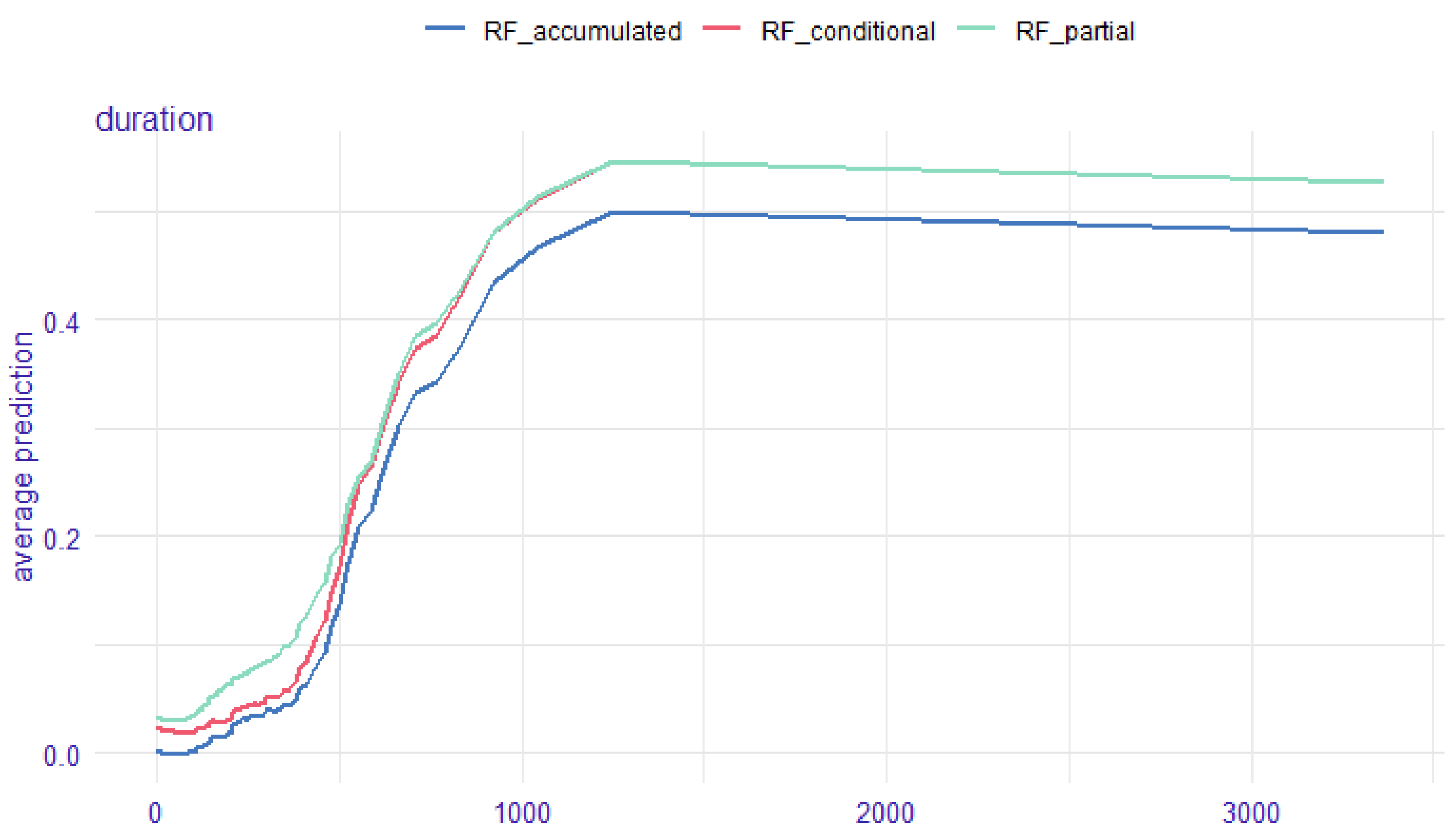
Accumalted local Effect Profile

- Based on the conditional distribution of the features – differences in predictions instead of averages



3. HOW CAN I USE XAI IN BUSINESS? (9/16)

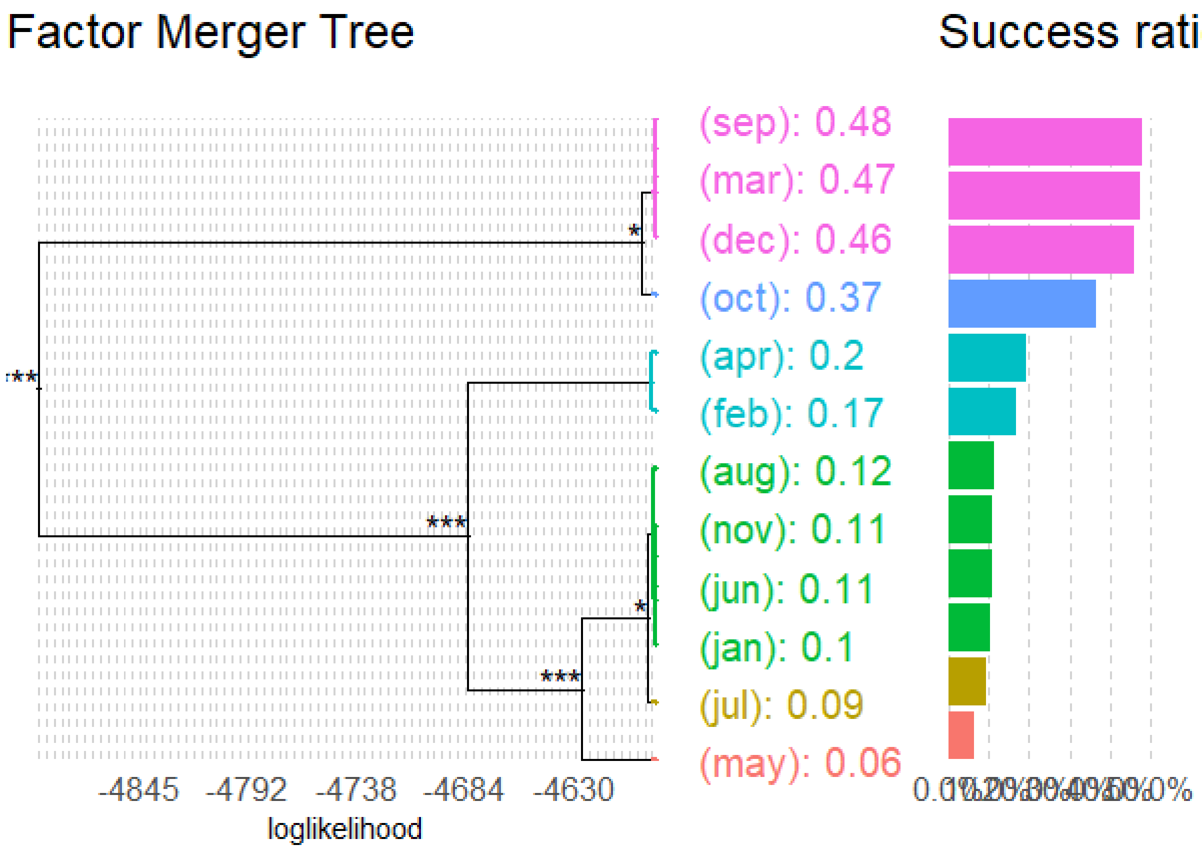
PDP, CDP and ALE plots



3. HOW CAN I USE XAI IN BUSINESS? (10/16)

PDP for Factors and factorMerger

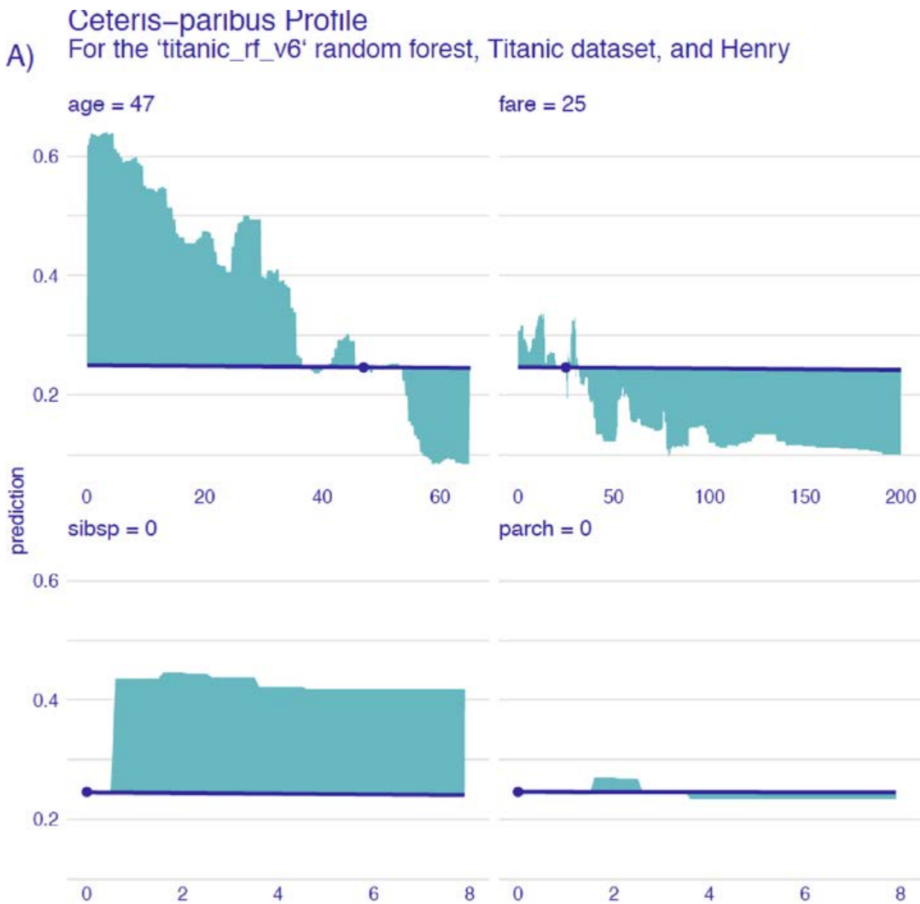
- Tool merging groups of categorical data into group o similar response based on LRT (likelihood ratio test)



3. HOW CAN I USE XAI IN BUSINESS? (11/16)

Oscilations

- Idea: the larger influence of indepvar on prediction, the larger fluctuations along CP profile.



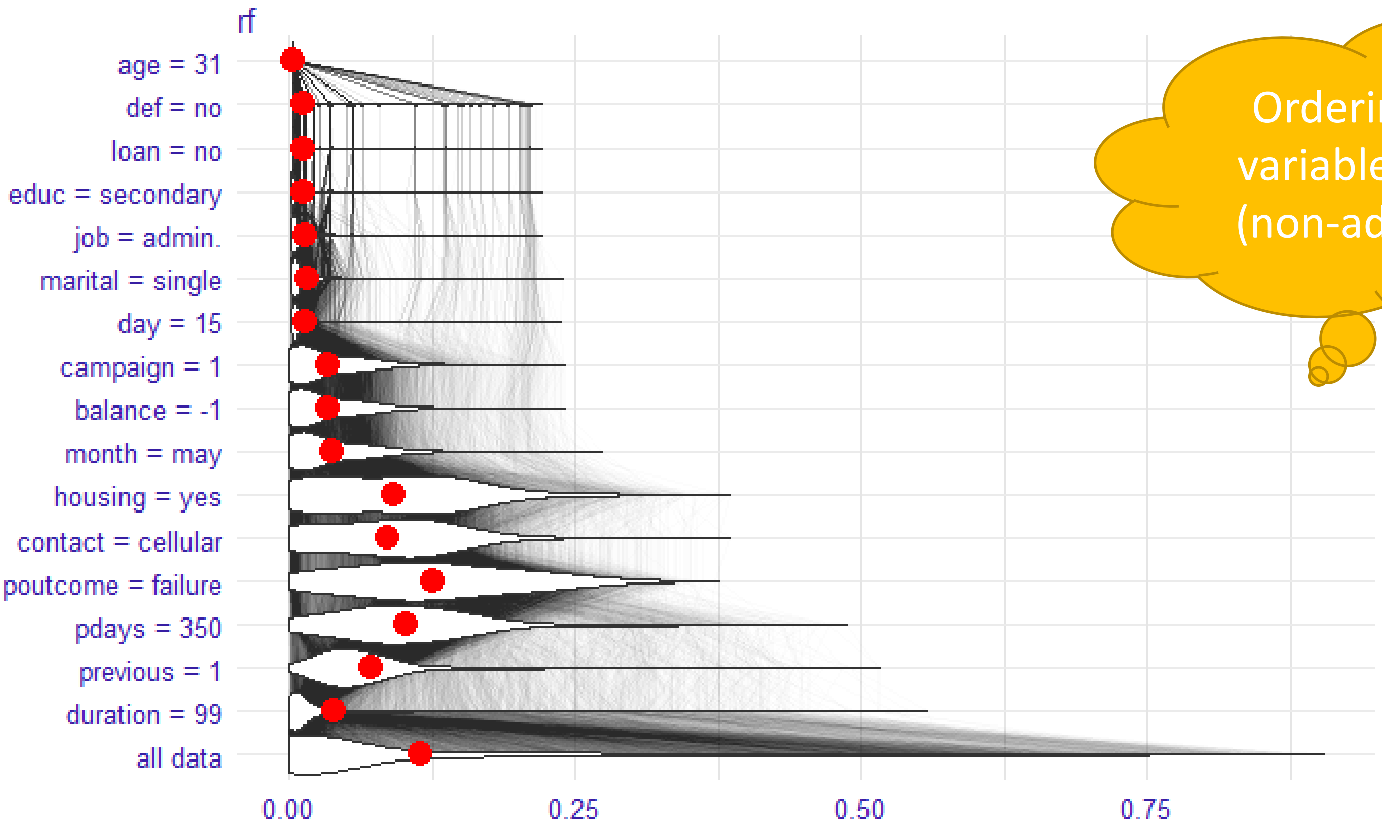
Pros:
- easy to interpret and understand.
- easily extendable to two or more variables.

Cons:
- Not working well if collinearity (based on CP)
- do not sum up to the instance prediction

3. HOW CAN I USE XAI IN BUSINESS? (12/16)

Break Down

- Idea: Decomposition of a prediction to baseline (average for a model) prediction and contribution of all variables separately



Ordering dependent variable contributions (non-additive models)

Avg. response for a model given duration = 99

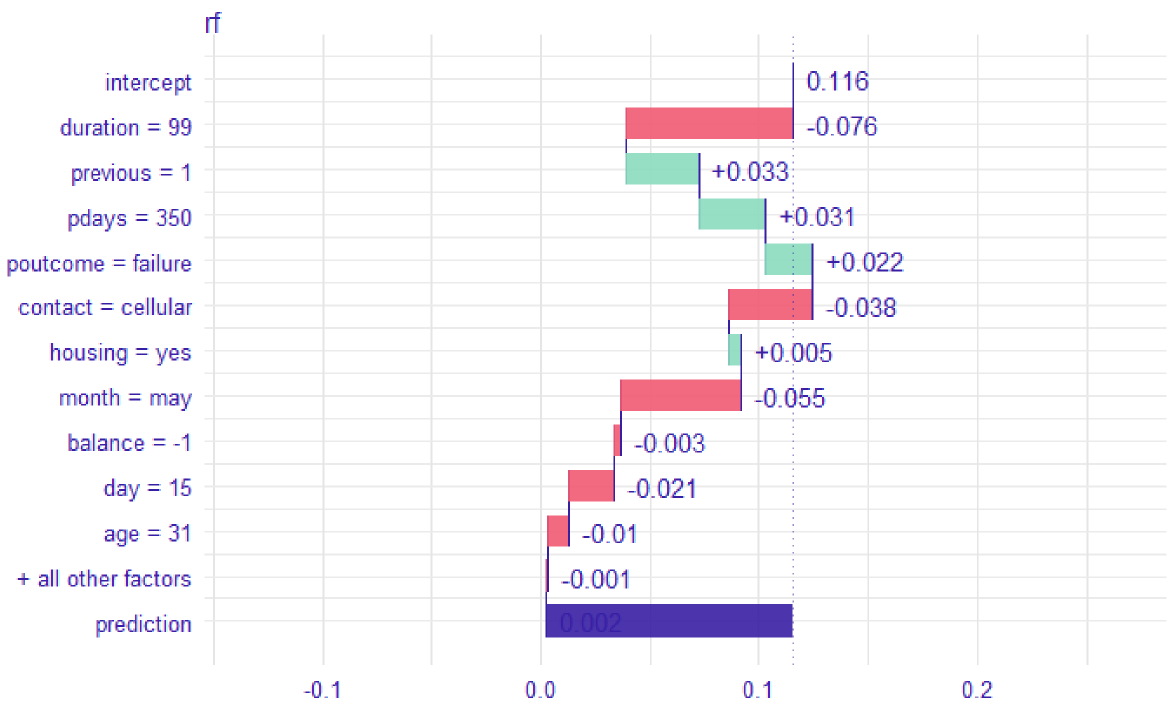
Avg. response for a mode

3. HOW CAN I USE XAI IN BUSINESS? (13/16)

Break Down – ordering dependency

Solutions:

- Ordering wrt local variable importance (BreakDown)
- Interaction identification (iBreakDown)
- Average from all paths (Shapley values)



Pros

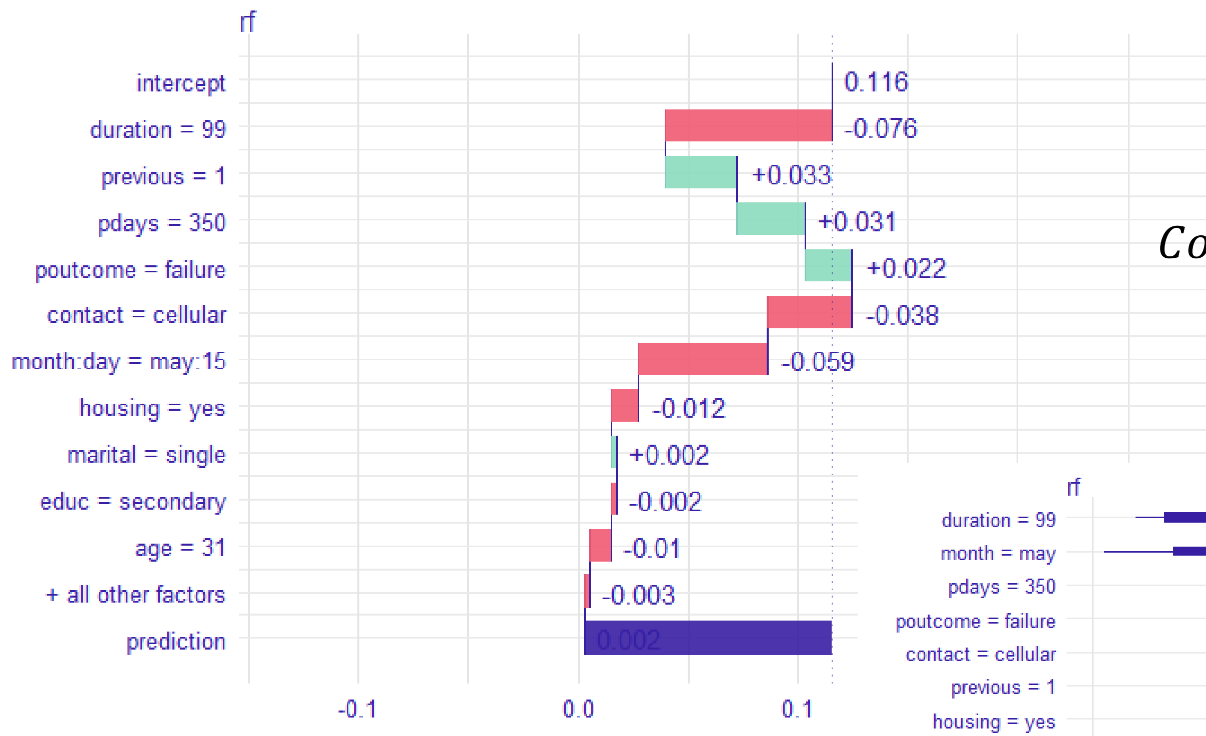
- easy to understand
- compact
- model agnostic
- complexity of Break Down Algorithm is linear in respect to the number of variables.

Cons

- selection of the ordering based on scores is subjective.
- different orderings may lead to different contributions.
- for non-additive model showing only additive contributions may be misleading.

3. HOW CAN I USE XAI IN BUSINESS? (14/16)

iBreakDown and Shapley values

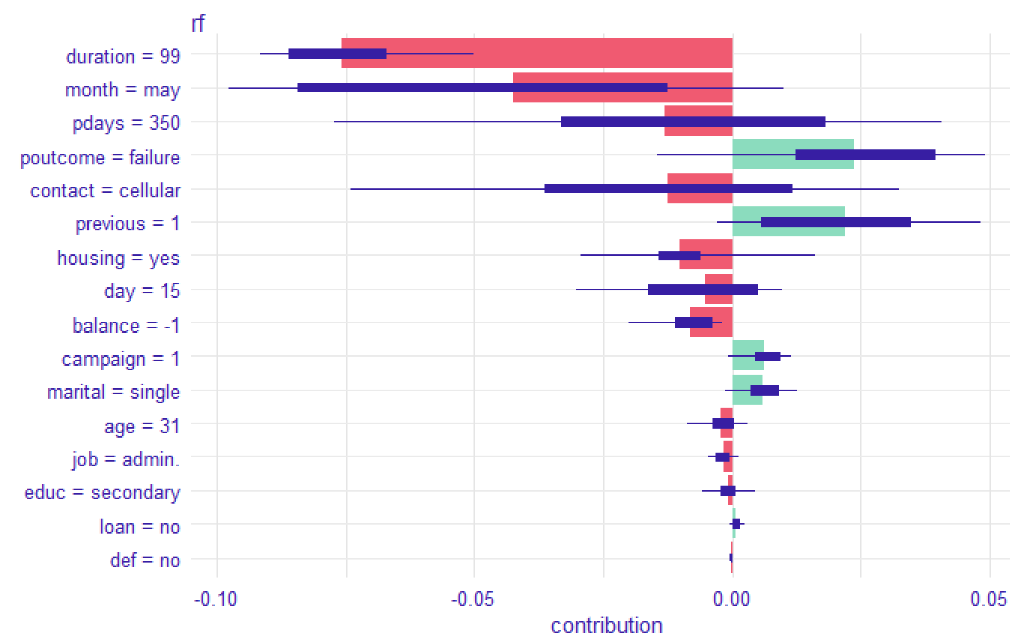


IDEA

$$Cont\ (x, y)\ \neq\ Cont(x) + Cont(y)$$

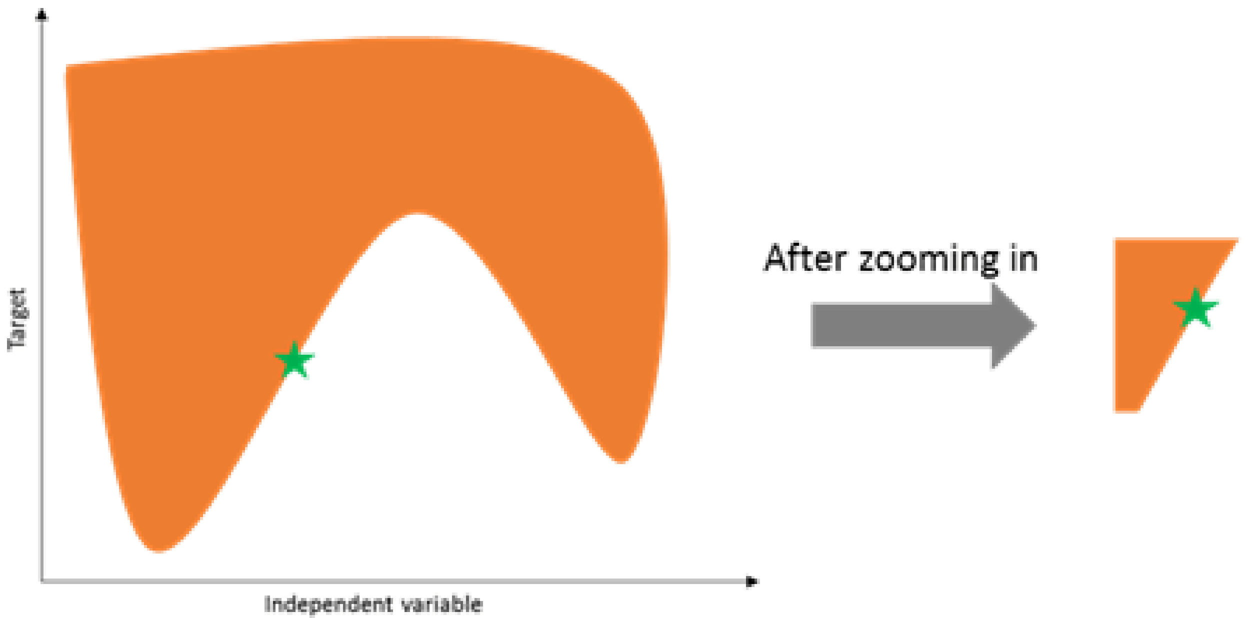
https://arxiv.org/abs/1903.11420

Averaging over all (large number of) paths



LIME

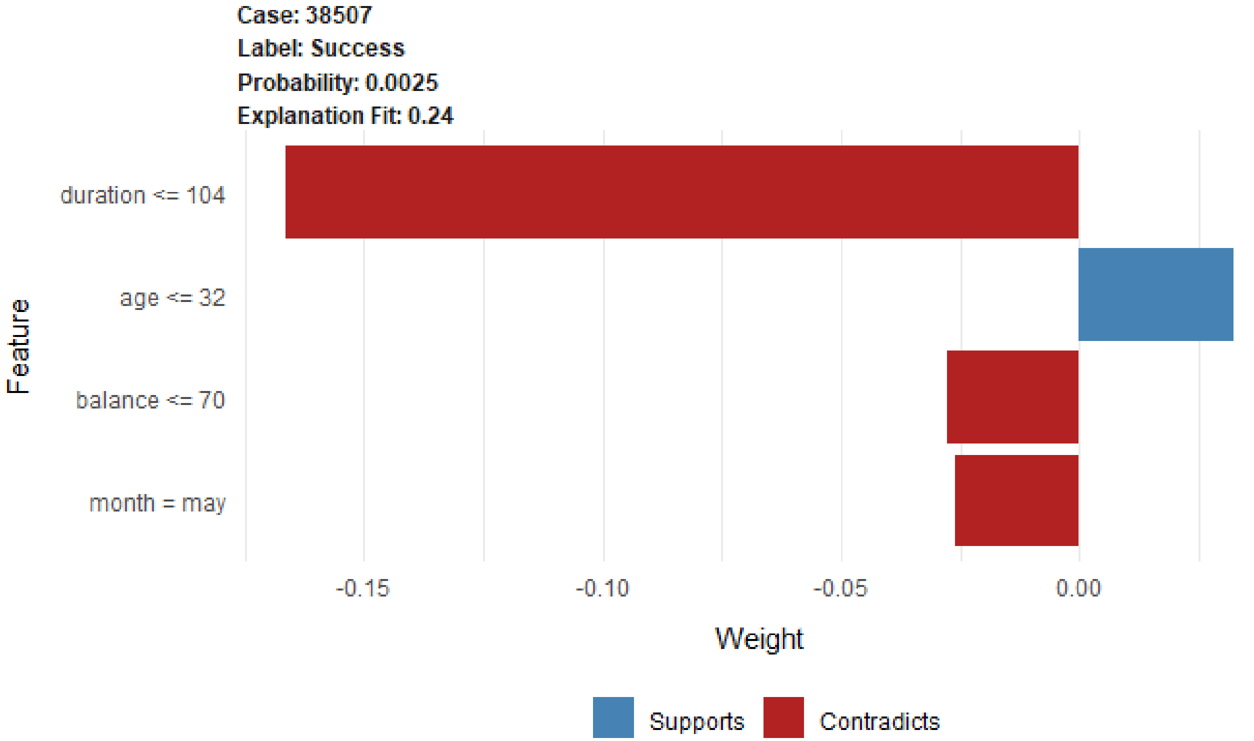
Local Interpretable Model-Agnostic Explanations. The key idea behind this method is to locally approximate a black-box model by a sparse local glass-box surrogate model, which is easier to interpret.



https://towardsdatascience.com/lime-explaining-predictions-of-machine-learning-models-1-2-1802d56addf9

3. HOW CAN I USE XAI IN BUSINESS? (16/16)

LIME



Pros

- Easy intuition
- Spare explanations
- Can be applied to high dimensional models.

Cons

- For continuous variables and tabular data it is not that easy to find interpretable representations.
- Explaining Phenomen by BB, and BB by WB.

https://github.com/pbiecek/xai_resources/blob/master/README.md#tools