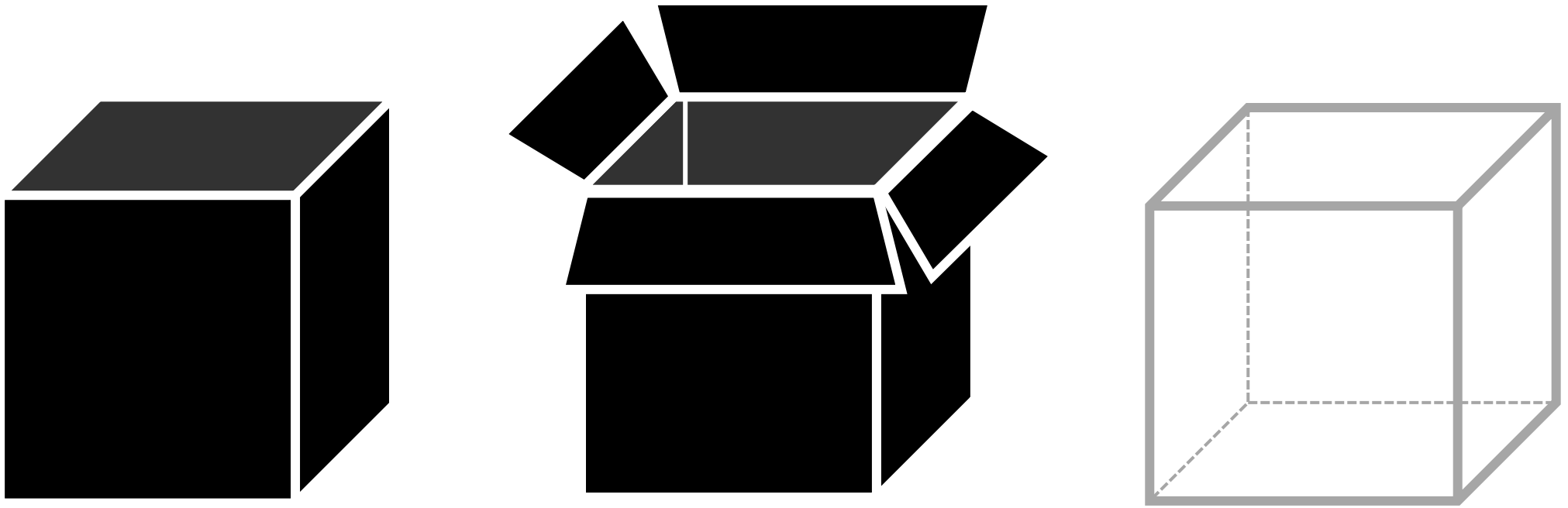
What’s inside the box?

An introduction to explainability in Finance

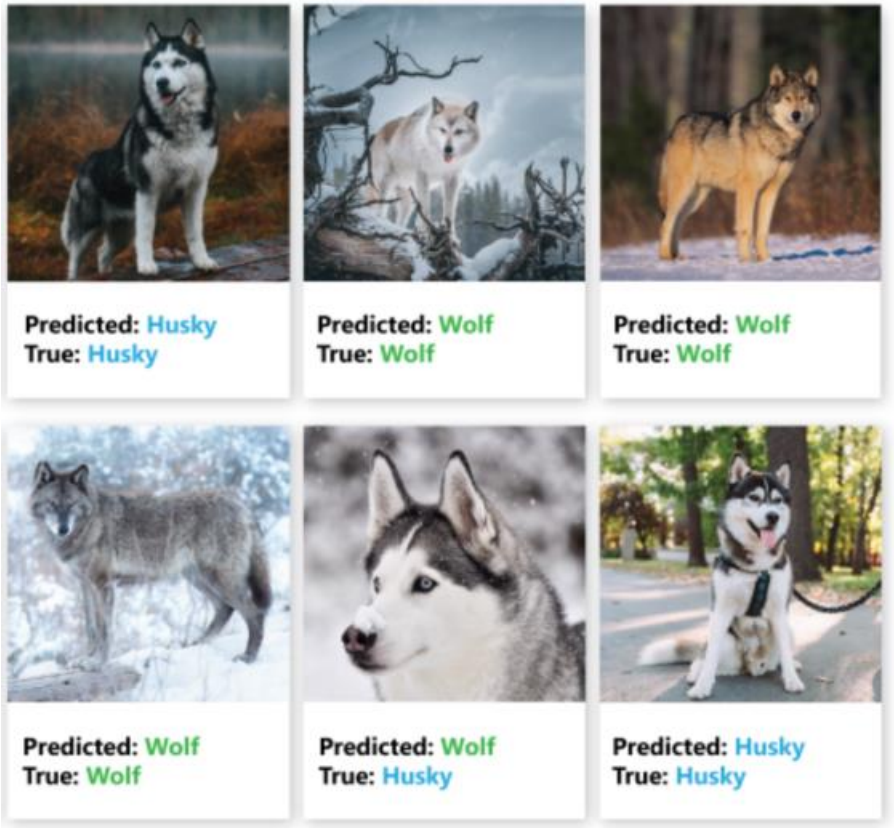
Magdalena Zych and Petra Gibcus

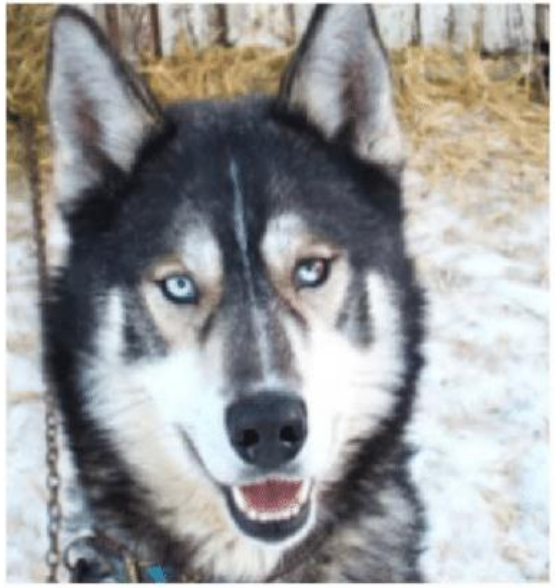
From black box to glass box



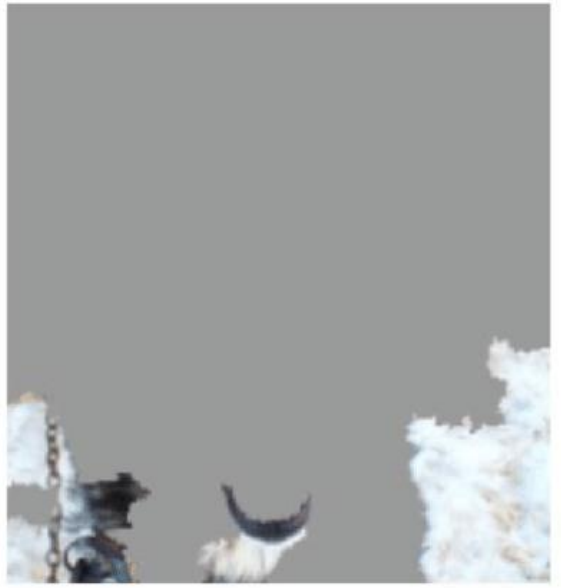
So, what does this look like in a machine learning model?

Would you trust this model?



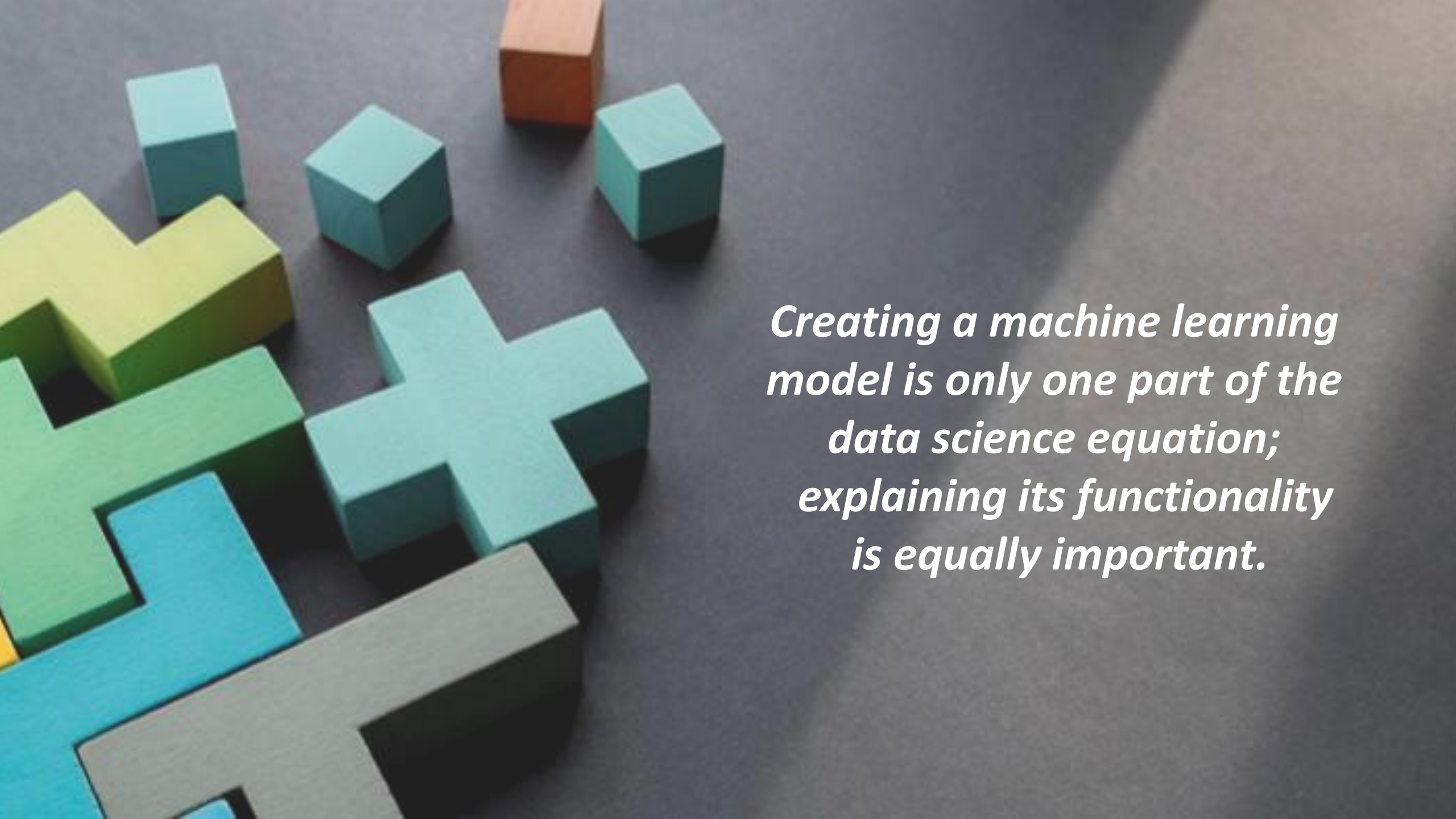


(a) Husky classified as wolf



(b) Explanation

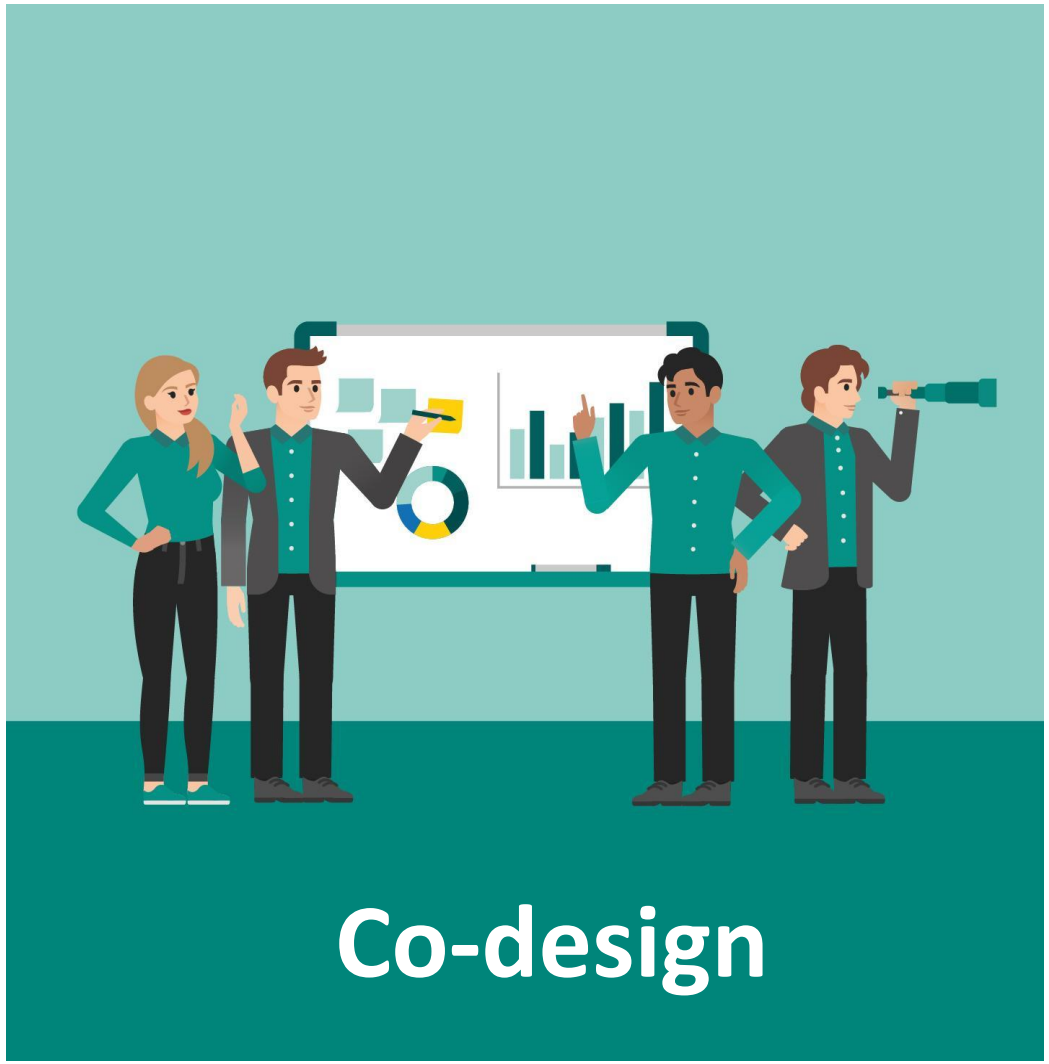
Besse, Philippe & Castets-Renard, Céline & Garivier, Aurélien & Loubes, Jean-Michel. (2018). Can Everyday AI be Ethical? Machine Learning Algorithm Fairness (english version). 10.13140/RG.2.2.22973.31207.
and https://carpentries-incubator.github.io/data-science-ai-senior-researchers/05-Problems-with-AI/index.html


Creating a machine learning model is only one part of the data science equation; explaining its functionality is equally important.

Why are we discussing this if our field is finance?

Importance of Explainable AI in finance

- Actionable insights
- Bias detection
- Regulatory compliance
- Trust predictions



Techniques for Explainable AI

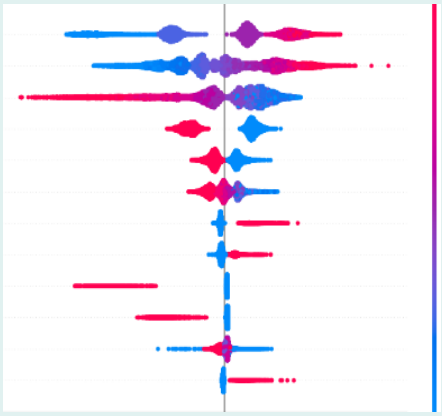
Model agnostic tools

LIME

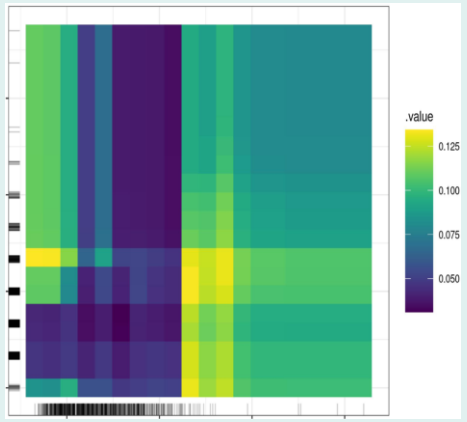
SHAP

Visualization tools

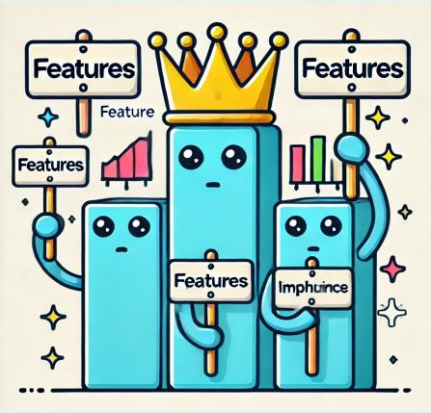
SHAP DASHBOARD



PARTIAL DEPENDENCE PLOT*



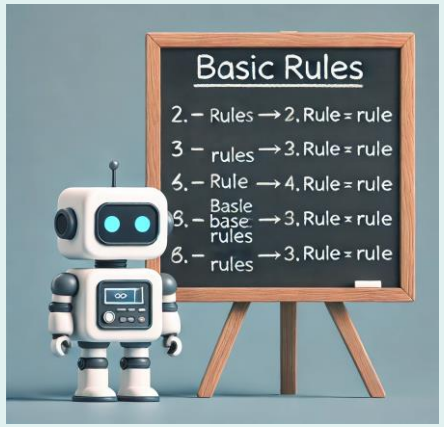
FEATURE IMPORTANCE PLOTS



Intrinsic methods

DECISION TREE

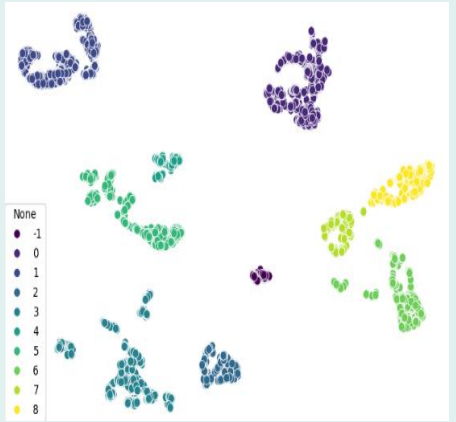
RULE-BASED SYSTEMS



SHAP clustering

UMAP

HBDSCAN



SKOPERULES

* Source: https://christophm.github.io/interpretable-ml-book/pdp.html

SHAPLEY
ADDITIVE
EXPLANATION

A Unified Way to Understand Any Model

SHAP? What is it?

SHAP is a method based on game theory that computes features contribution to predictions in ML model on both global and local level.

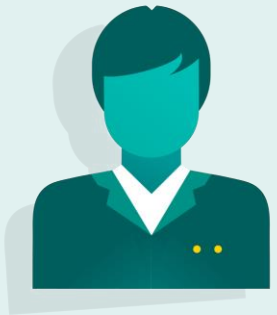
Coalition values*

C_{12} = 10,000
C_1 = 7,500
C_2 = 5,000
C_0 = 0

$C_{12} - C_2$ = 5,000
$C_1 - C_0$ = 7,500

(5,000 + 7,500) / 2
= € 6,250

Marginal contribution
The increase in a coalition's value due to a player joining that coalition



$C_{12} - C_1$ = 2,500
$C_2 - C_0$ = 5,000

(2,500 + 5,000) / 2
= € 3,750

Shapley value = expected marginal contribution

* Example based on https://www.youtube.com/watch?v=UJeu29wq7d0&list=PLqDyyww9y-1SJgMw92x90qPYpHgahDLIK&index=4

Workshop on Explainable AI
Mortgage churn use case

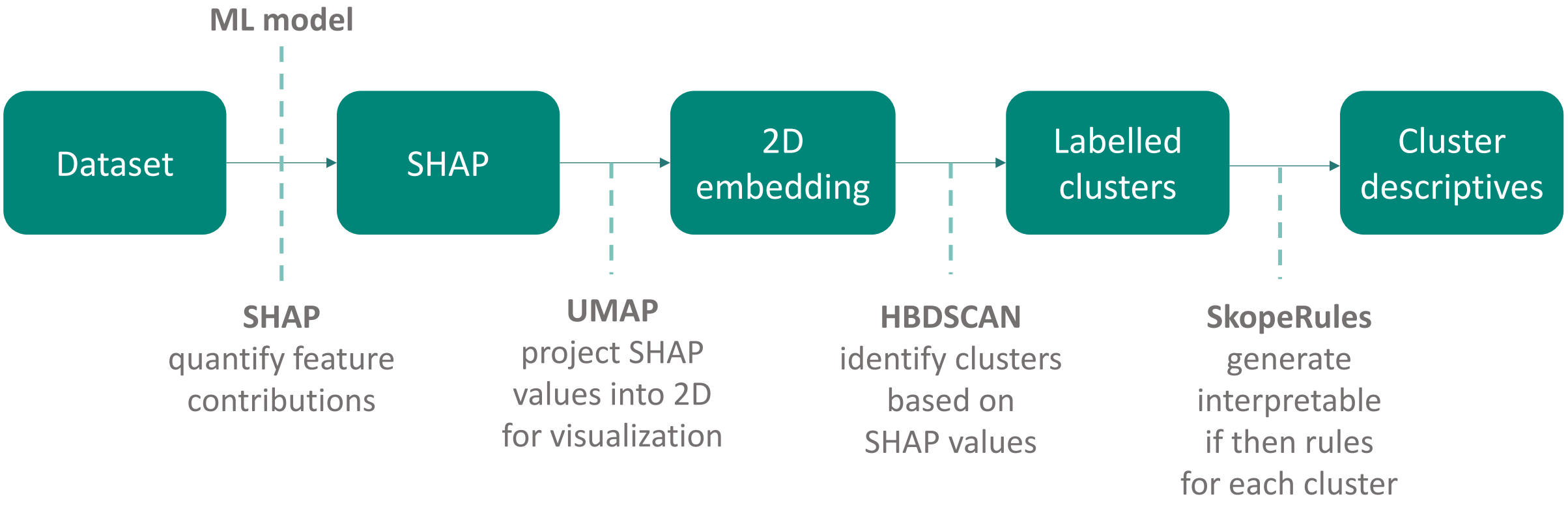
Use case mortgage churn

Customer mortgage churn prediction

Understand why and which customers leave

Improve customer mortgage retention

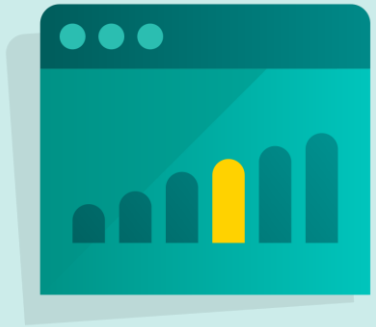
Workflow


ML model
Dataset
SHAP
2D embedding
Labelled clusters
Cluster descriptives
SHAP
quantify feature contributions
UMAP
project SHAP values into 2D for visualization
HBDSCAN
identify clusters based on SHAP values
SkopeRules
generate interpretable if then rules for each cluster


Exercise
1
2
3

Dataset use case mortgage churn

- Generated data*
- No ABN AMRO customers
- Not representative for ABN AMRO



- All data is ordered
- All data is cleaned
- Ordinal and continuous variables

- Target: churn
- Balanced data

* Credits: dataset is adapted based on: https://archive.ics.uci.edu/dataset/34/diabetes

Grab the dataset from the repo

Get the notebook

Code along!!!

Exercise 1: Introduction to SHAP visuals

20 minutes

1.1 *Global SHAP*

Generate two SHAP summary plots

- one for a model trained with the leaky feature (model_b)
- another without it (model)

1.2 *Local SHAP*

Fill in samples in the waterfall plot.

1.3 *Comparing SHAP explanations with different background samples* ***(BONUS)***

Based on EDA choose appropriate activity level that would describe a subset of people more likely to churn based on their activity

Exercise 2: SHAP and feature interactions

15 minutes

2.1 *SHAP and clusters dendrogram*

Make a dependence plot between the top two clustered features.

2.2 *SHAP and highly correlated features*

- Make a dependence plot between *age* and randomly generated feature *n_correlated_noise age*
- Experiment with different feature correlations

Exercise 3: SHAP clustering

15 minutes

3.1 *Calculate SHAP values*

Run the cell in the notebook.

3**.2** *UMAP dimensionality reduction*

Run the cell in the notebook.

3.3 *HBDSCAN clustering*

Are there clusters that contain solely specific values? Check it out!

3.4 *SkopeRules to extract meaningful insights* ***(BONUS)***

Run the cell in the notebook. Can you explain the results?

Learnings SHAP

ADVANTAGES

- Model agnostic
- Local and global explainability for feature importance
- Capturing feature interactions
- Visualising results

DISADVANTAGES

- Computationally intensive
- Challenging: interpretation of the results
- Not to be used as causal or counterfactual explanations
- Sensitive to correlation in data

KEY TAKE AWAYS

- Explainability is the key in data science projects
- Keeping stakeholders in the loop is essential
- EDA should be reflected in model interpretations
- SHAP is not the holy grail of explainability

Thanks to…