Explore, Explain, and Debug

aka Interpretable Machine Learning

Przemysław Biecek

Use case
FIFA 2019

Dataset

FIFA 19 complete player dataset

18k+ FIFA 19 players, ~90 attributes extracted from the latest FIFA database

Karan Gadiya • updated a year ago

1640

Data Kernels (304) Discussion (21) Activity Metadata Download (9 MB) New Notebook

Your Dataset download has started.
Show your appreciation with an upvote

1640

Usability 10.0 **License** CC BY-NC-SA 4.0 **Tags** data visualization, feature engineering, random forest, sports, regression analysis

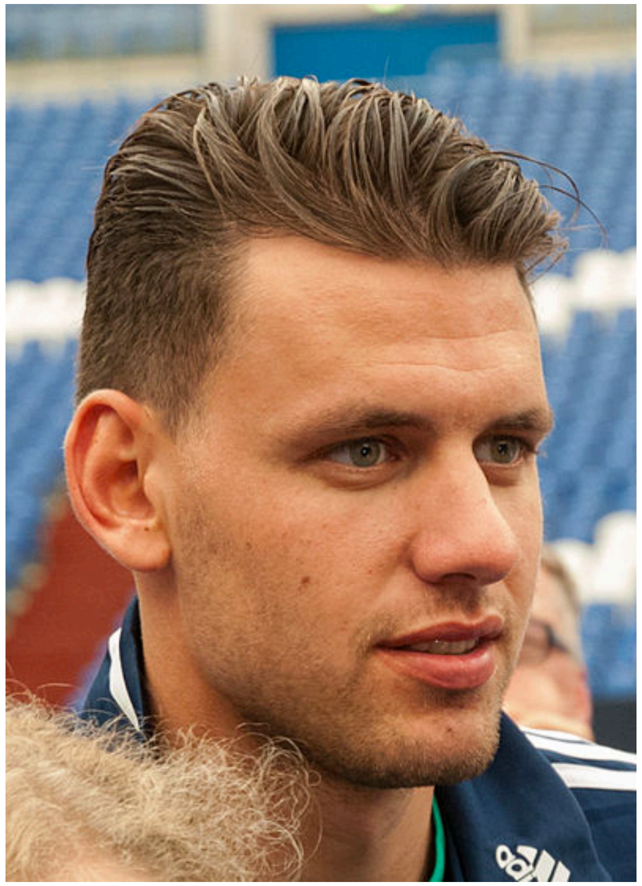
Description

Context

Football analytics

```
library("gbm")
fifa_gbm <- gbm(ValueNum~.,
        data              = fifa19,
        n.trees           = 250,
        interaction.depth = 3)
```

Source: https://www.kaggle.com/karangadiya/fifa19/data

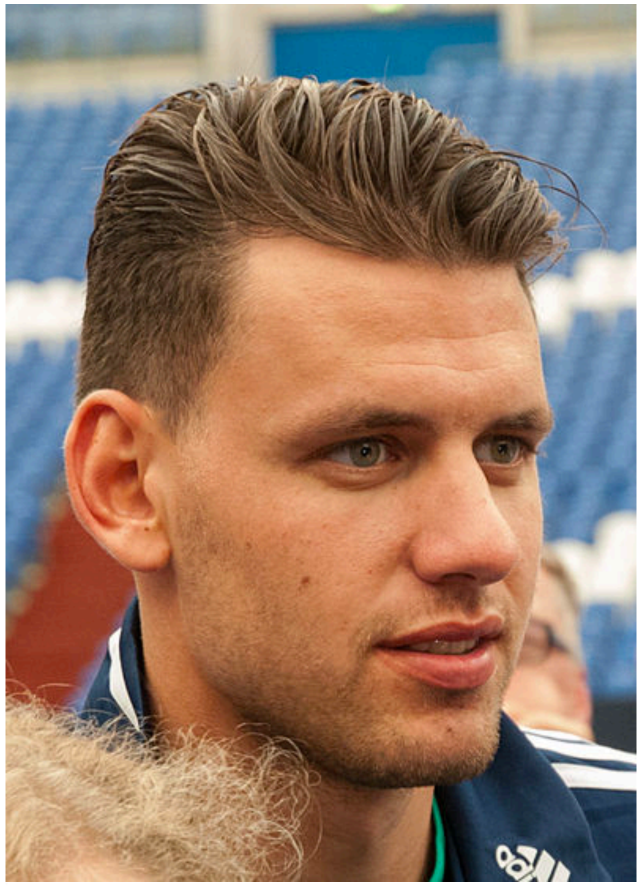


wikipedia

Full name	Ádám Csaba Szalai
Date of birth	9 December 1987
Place of birth	Budapest, Hungary
Playing position	Striker

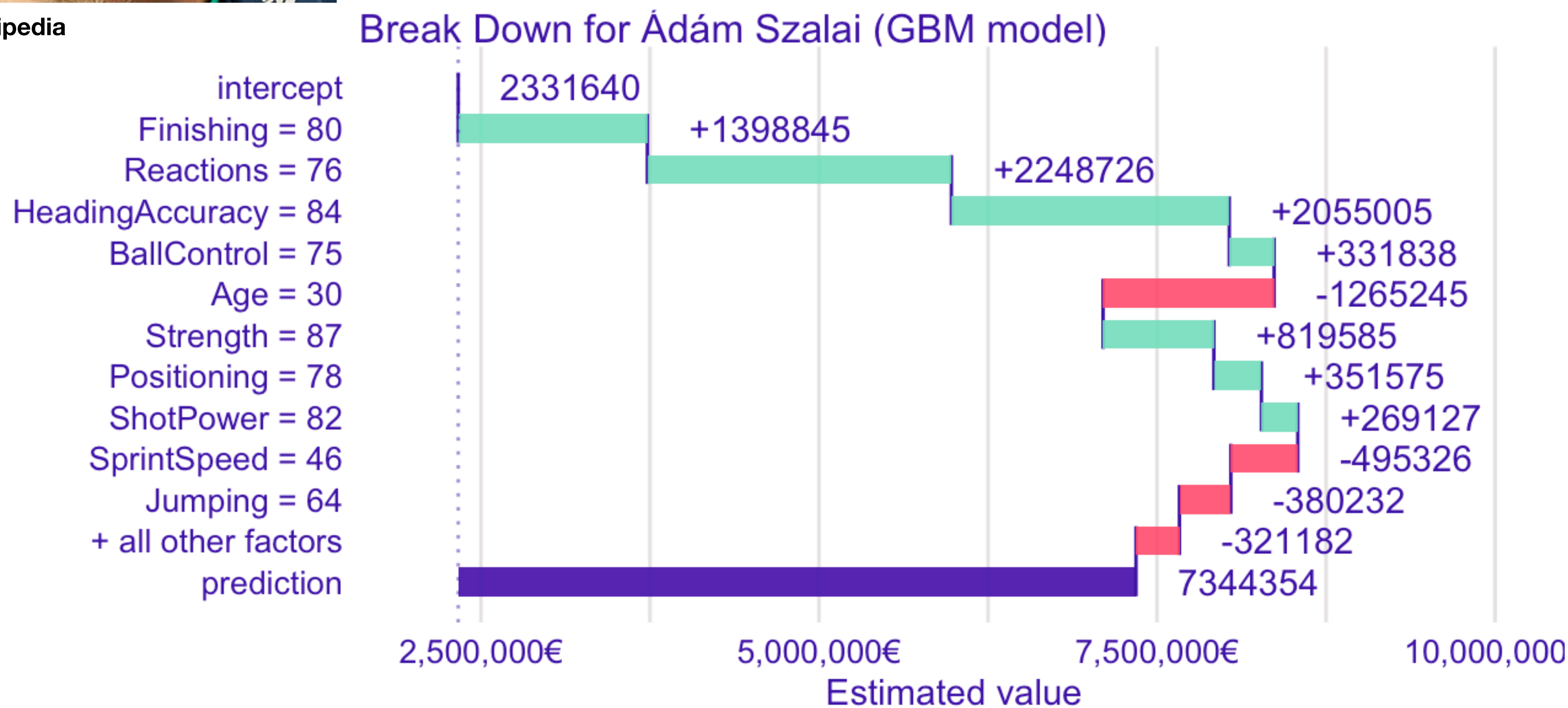
Prediction from GBM model: 7 344 354 EUR

How? Why? Really?



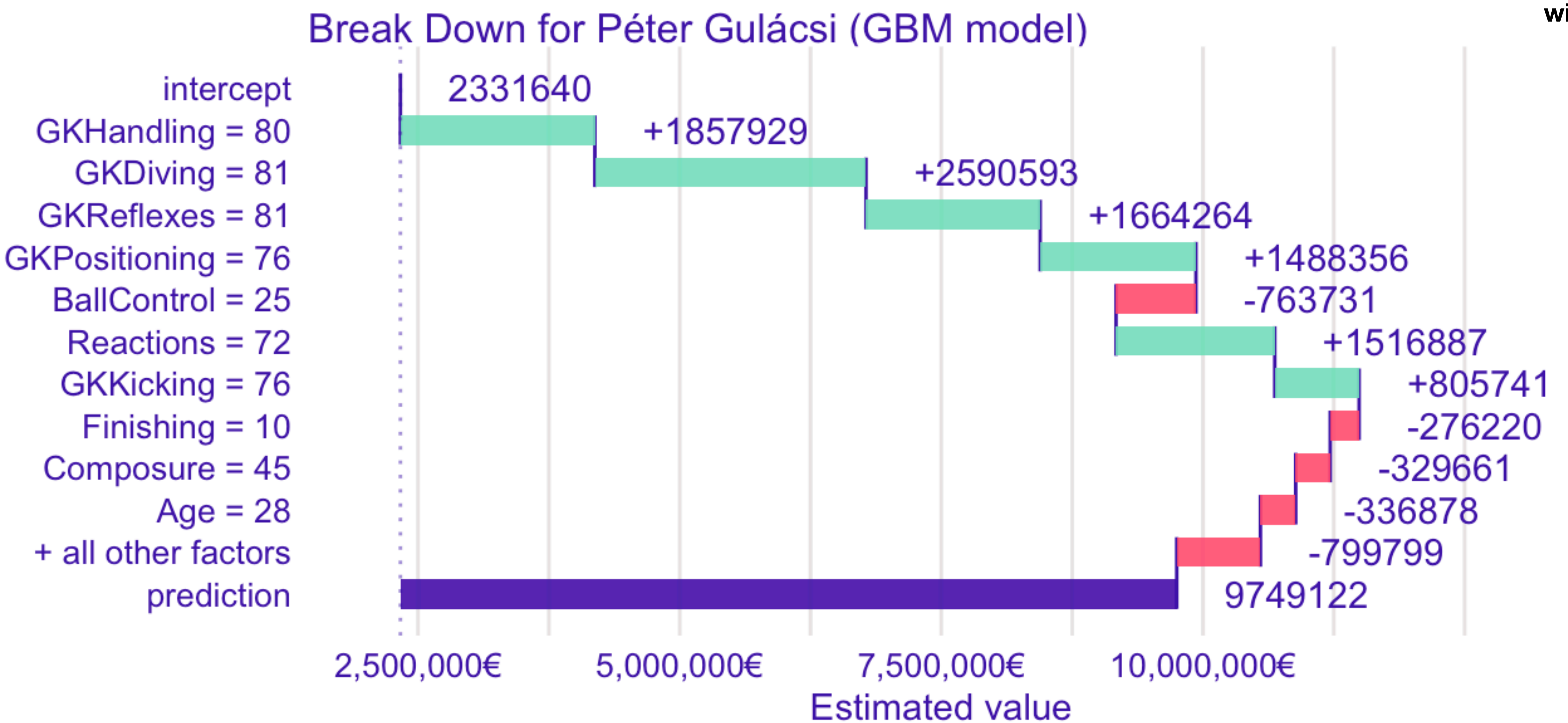
wikipedia

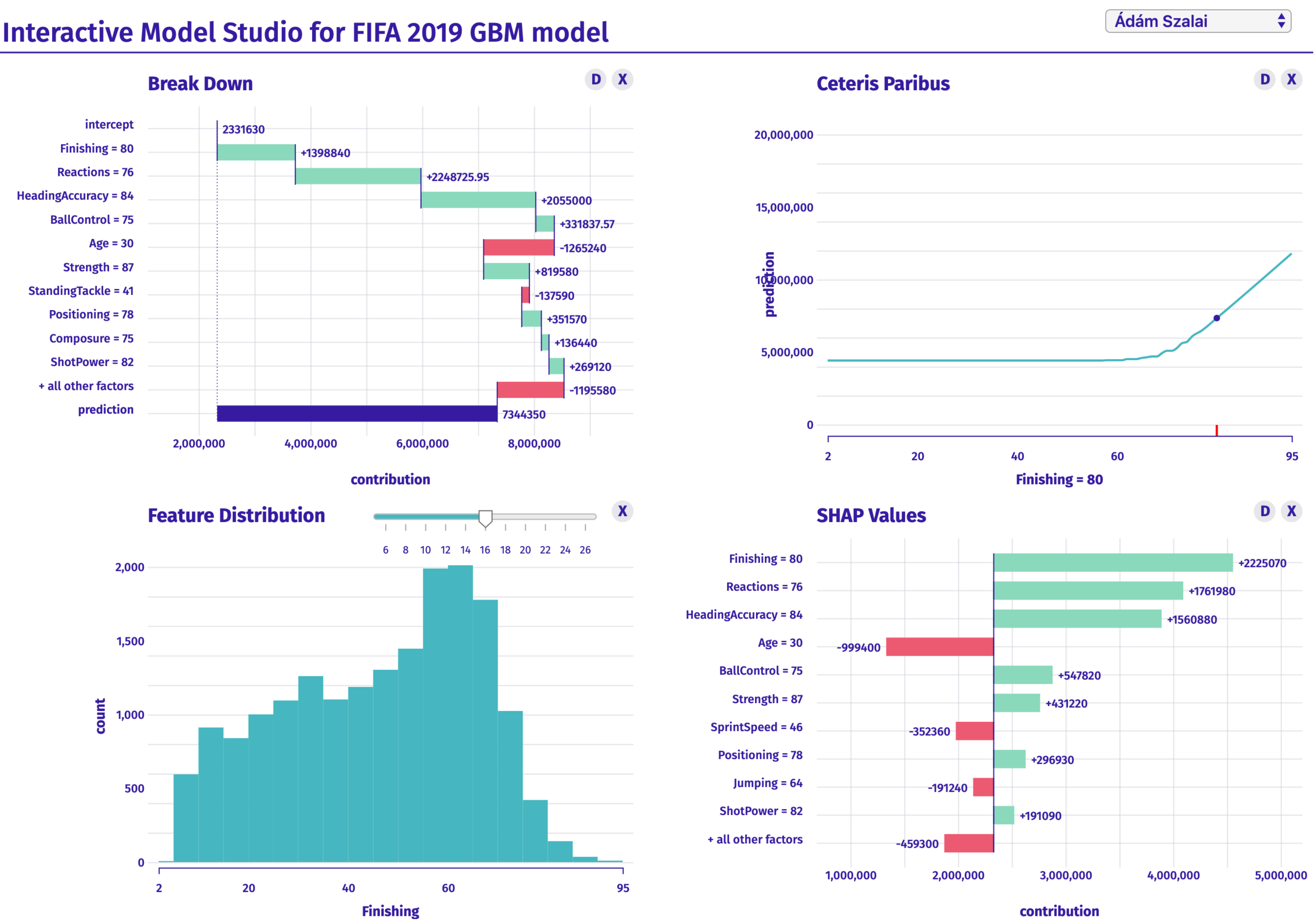
Full name	Ádám Csaba Szalai
Date of birth	9 December 1987
Place of birth	Budapest, Hungary
Playing position	Striker



Full name	Péter Gulácsi
Date of birth	6 May 1990
Place of birth	Budapest, Hungary
Playing position	Goalkeeper

wikipedia





https://pbiecek.github.io/explainFIFA19/

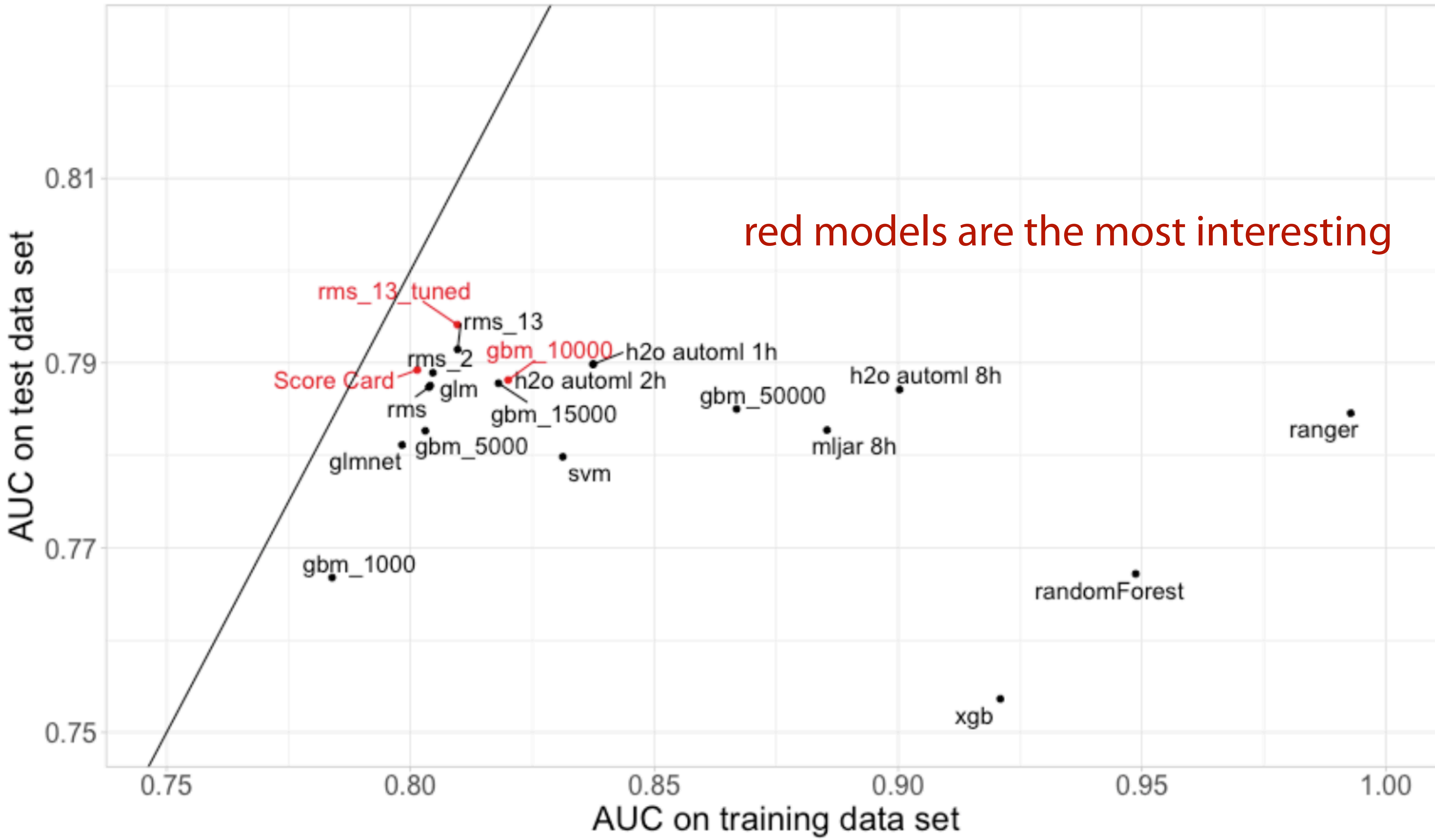
Use case
Credit Scoring

Data set for study: xML Challenge by FICO

- Explainable Machine Learning Challenge by FICO (2019)
- Focus: Home Equity Line of Credit (HELOC) Dataset
- Customers requested a credit line in the range of $5,000 - $150,000
- Task is to predict whether they will repay their HELOC account within 2 years
- Number of observations: 2,615
- Variables: 23 covariates (mostly numeric) and 1 target variable (risk performance "good" or "bad")

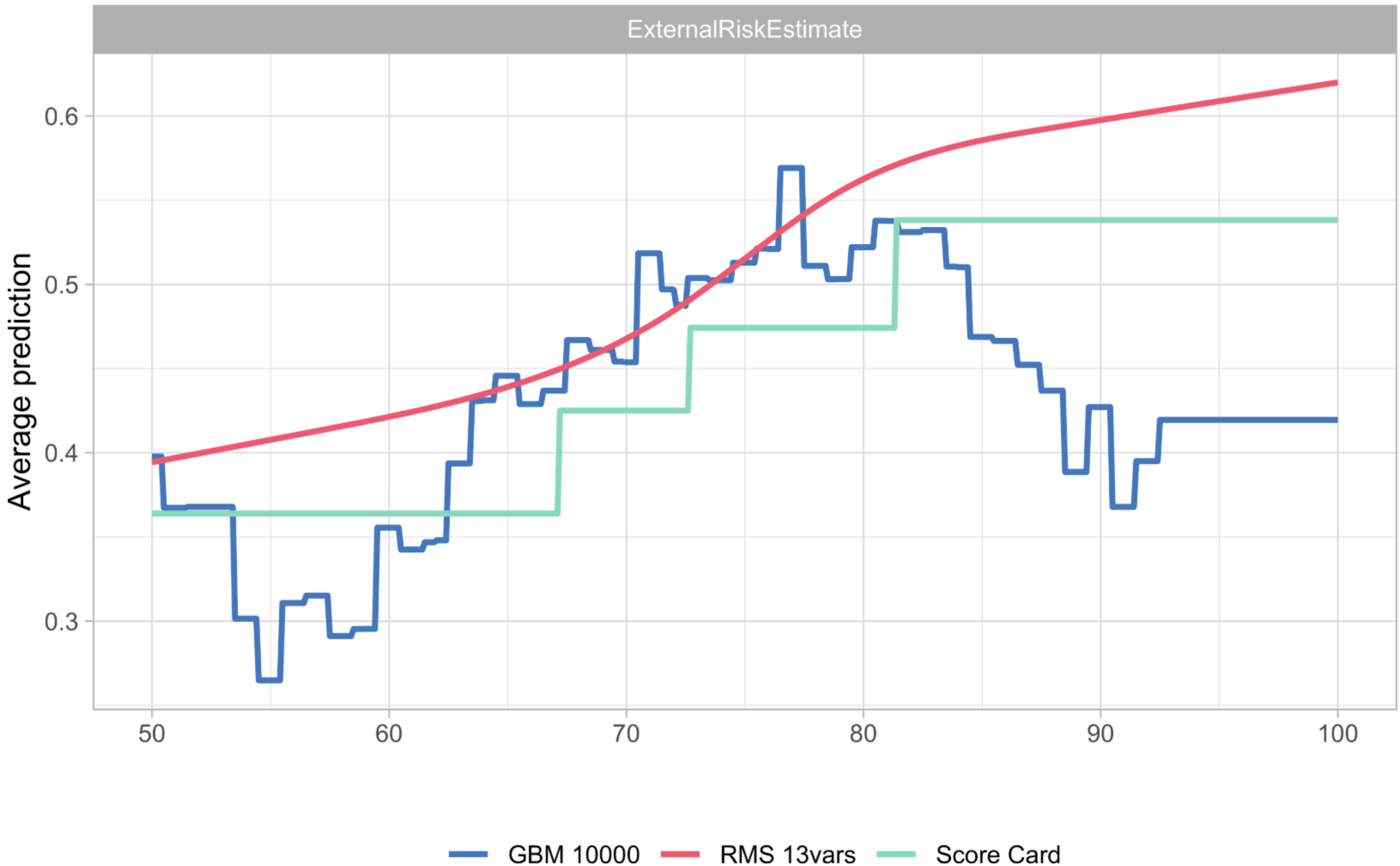
https://buecker.ms/slides/crccxvi/slides.html#1

Performance for selected modeling methods



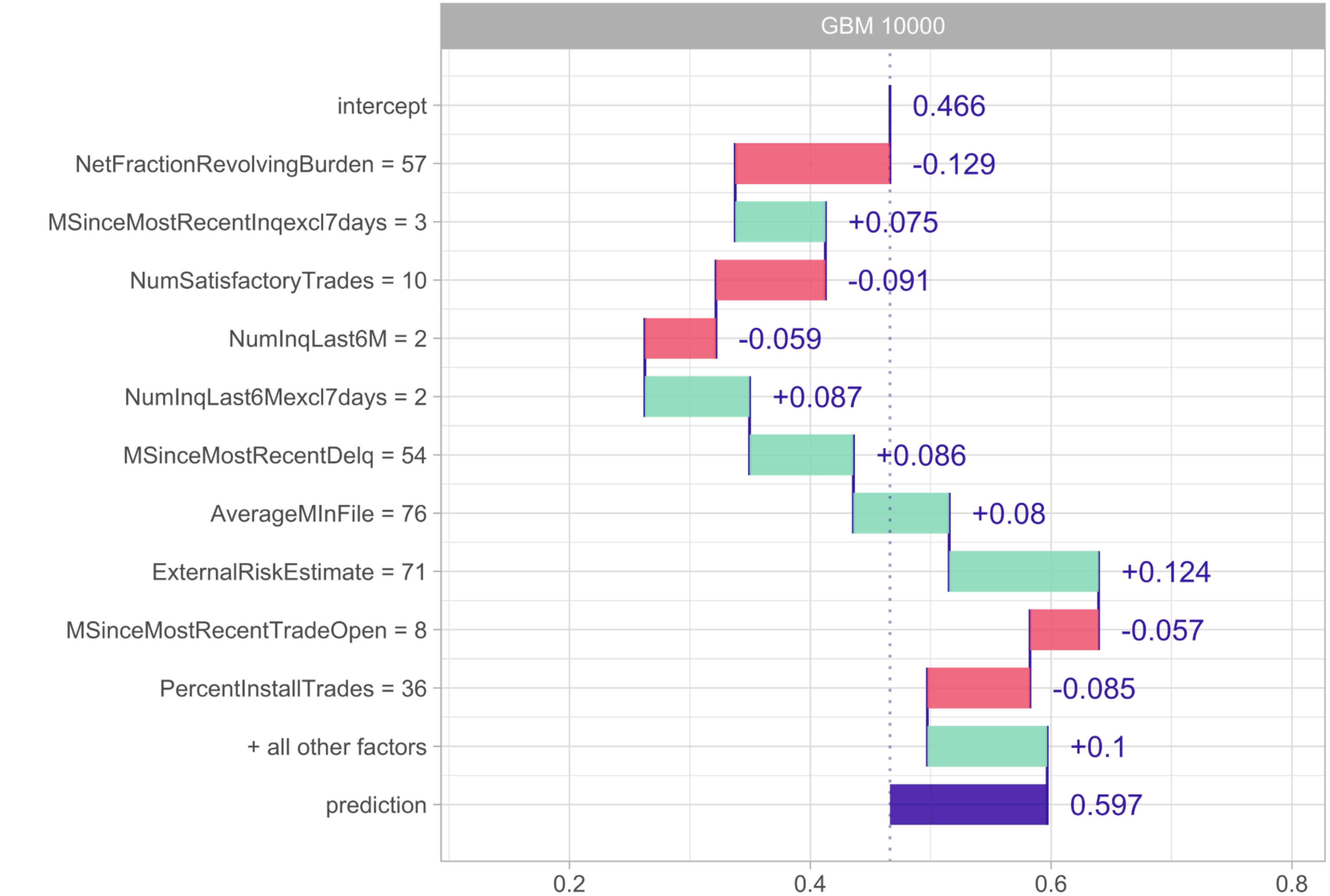
https://buecker.ms/slides/crccxvi/slides.html#1

Partial Dependency Plot for the most important feature



https://buecker.ms/slides/crccxvi/slides.html#1

Model agnostic: Variable contribution break down



- Such instance-lev explorations can be performed in a model-agnostic w
- Unfortunately, fo non-additive mod variable contributions depend on the ordering of varial

https://buecker.ms/slides/crccxvi/slides.html#1

FH MÜNSTER
University of Applied Sciences

MSB FB Wirtschaft
Münster School of Business

Transparency of Machine Learning Models in Credit Scoring

CRC Conference XVI

Michael Bücker, Gero Szepannek, Przemyslaw Biecek,
Alicja Gosiewska and Mateusz Staniak

28 August 2019

https://buecker.ms/slides/crccxvi/slides.html#1

Model specific / agnostic explainability

Model specific

+ Exploits the structure of a model
+ Allow for deep diagnostic

- Explanations cannot be easily compared between models
- You need to create an explainer for every possible model structure

Examples:

- randomForestExplainer
- EIX - lightgbm/xgboost explainer

Model agnostic

+ Explanations can be compared between different models
+ Can be used to any model (model is a black box)

- Explanations are (im most cases) approximations, they may be inaccurate or wrong
- It is easy to miss some quirks

Examples:

LIME / SHAP / Break Down / Partial Dependency Profiles / Permutational Feature Importance

Local / global explainability

Local explanations

+ Focused on a single observation
+ Good for dissecting of a single prediction
+ Good for debugging

– Feature effects may be different for different observations
– Tricky for correlated features

Examples:

- LIME / SHAP / Break Down
- Ceteris Paribus

Global explanations

+ Focused on a model
+ Good for global model summaries
+ Good for comparisons

– Local behaviour may be different
– For non-additive models they may be to simplistic

Examples:

- Feature Importance
- Partial Dependency Profiles

ModelOriented / DALEX

Unwatch 28 | Unstar 493 | Fork 77

Code | Issues 15 | Pull requests 0 | Projects 0 | Wiki | Security | Insights | Settings

Branch: master | DALEX / README.md | Find file | Copy path

85 lines (56 sloc) | 7.35 KB | Raw | Blame | History

Descriptive mAchine Learning EXplanations

build passing | coverage 66% | CRAN 0.4.9 | downloads 38K | DrWhy BackBone

Overview

The `DALEX` package (Descriptive mAchine Learning EXplanations) helps to understand how complex models are working. The main function `explain()` creates a wrapper around a predictive model. Wrapped models may then be explored and compared with a collection of local and global explainers. Recent developents from the area of Interpretable Machine Learning/eXplainable Artificial Intelligence.

The philosophy behind `DALEX` explanations is described in the Predictive Models: Explore, Explain, and Debug e-book. The `DALEX` package is a part of DrWhy.AI universe.

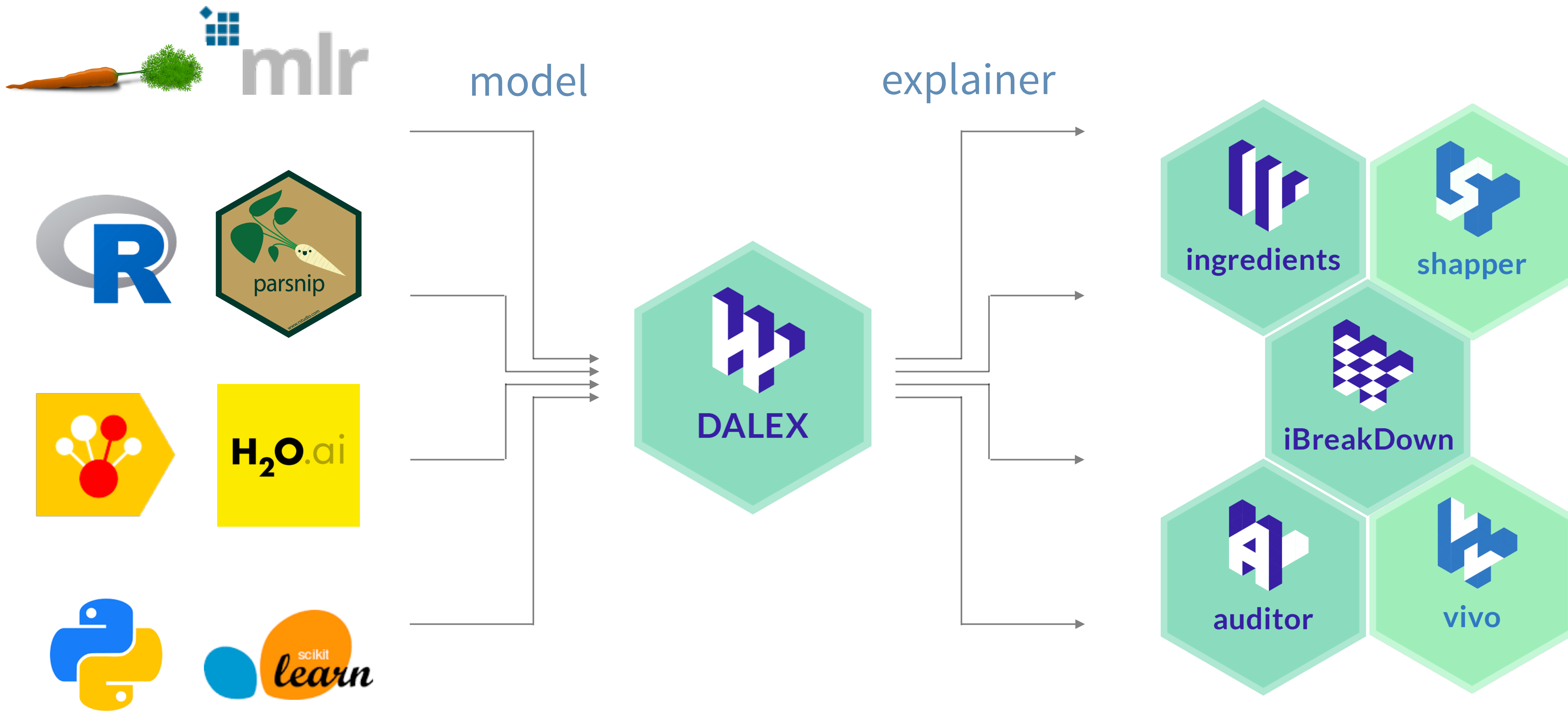
If you work with `scikitlearn`, `keras`, `H2O`, `mljar` or `mlr`, you may be interested in the DALEXtra package. It is an extension pack for `DALEX` with easy to use connectors to models created in these libraries.

https://github.com/ModelOriented/DALEX

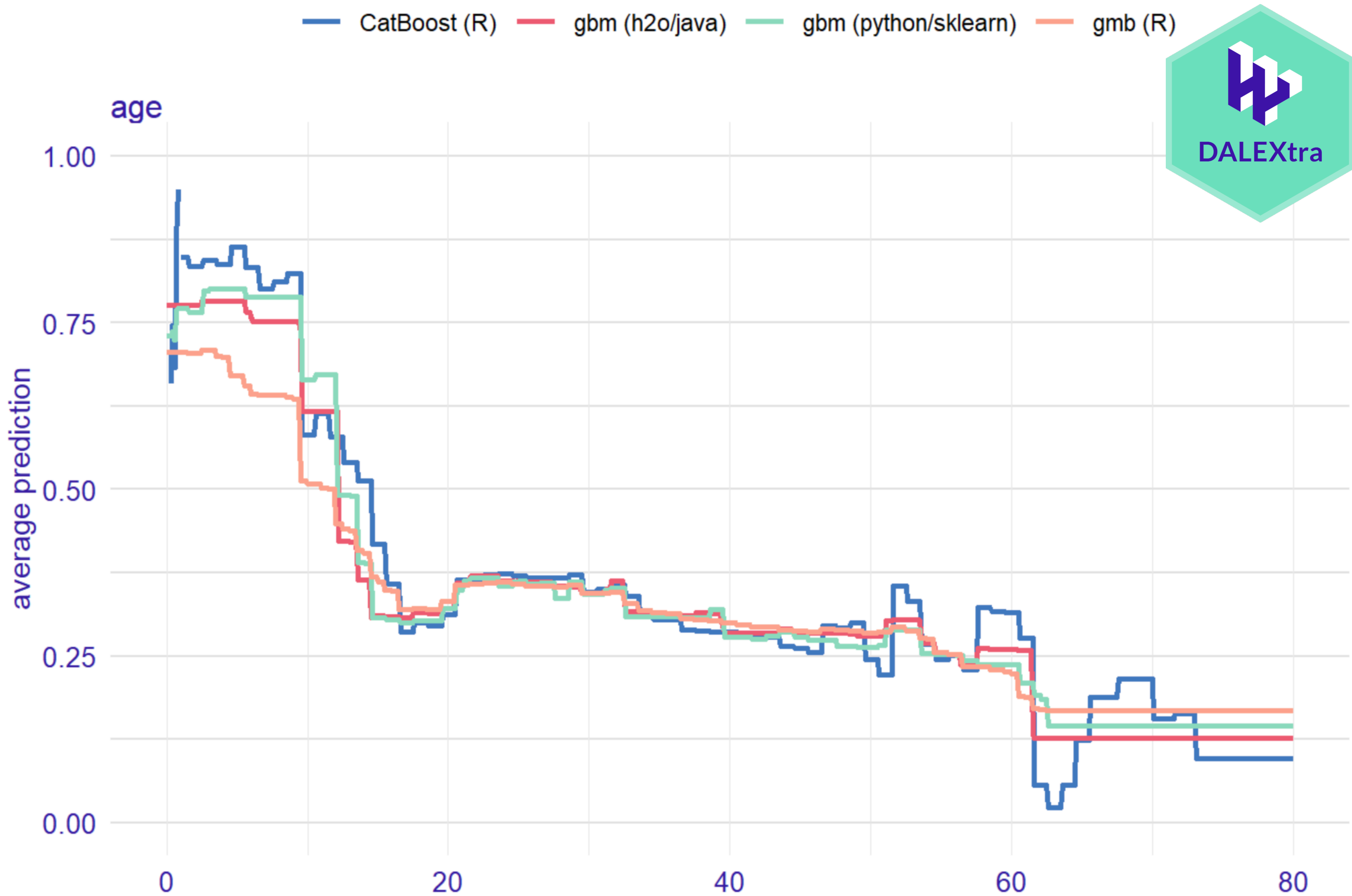
Make models
not wars

DALEX wraps a model available in R or python or through REST API into a standardised container.

Other tools for XAI may then work on the model despite its internal structure.



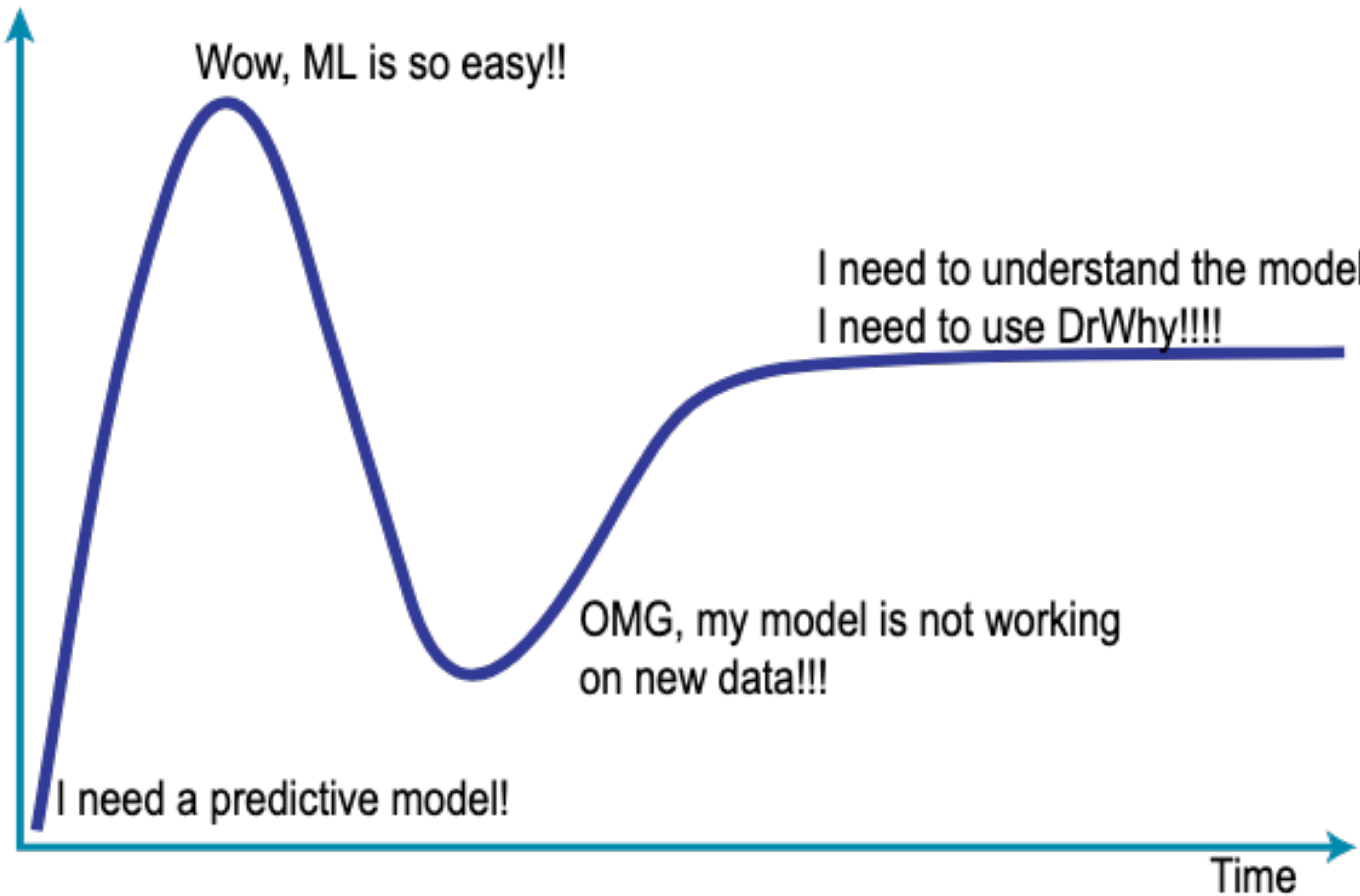
https://medium.com/@ModelOriented/xai-in-the-jungle-of-competing-frameworks-for-machine-learning-fa6e96a99644



https://medium.com/@ModelOriented/xai-in-the-jungle-of-competing-frameworks-for-machine-learning-fa6e96a99644

Why should we care?

Hype Cycle for Predictive Models



http://drwhy.ai

Thank you!