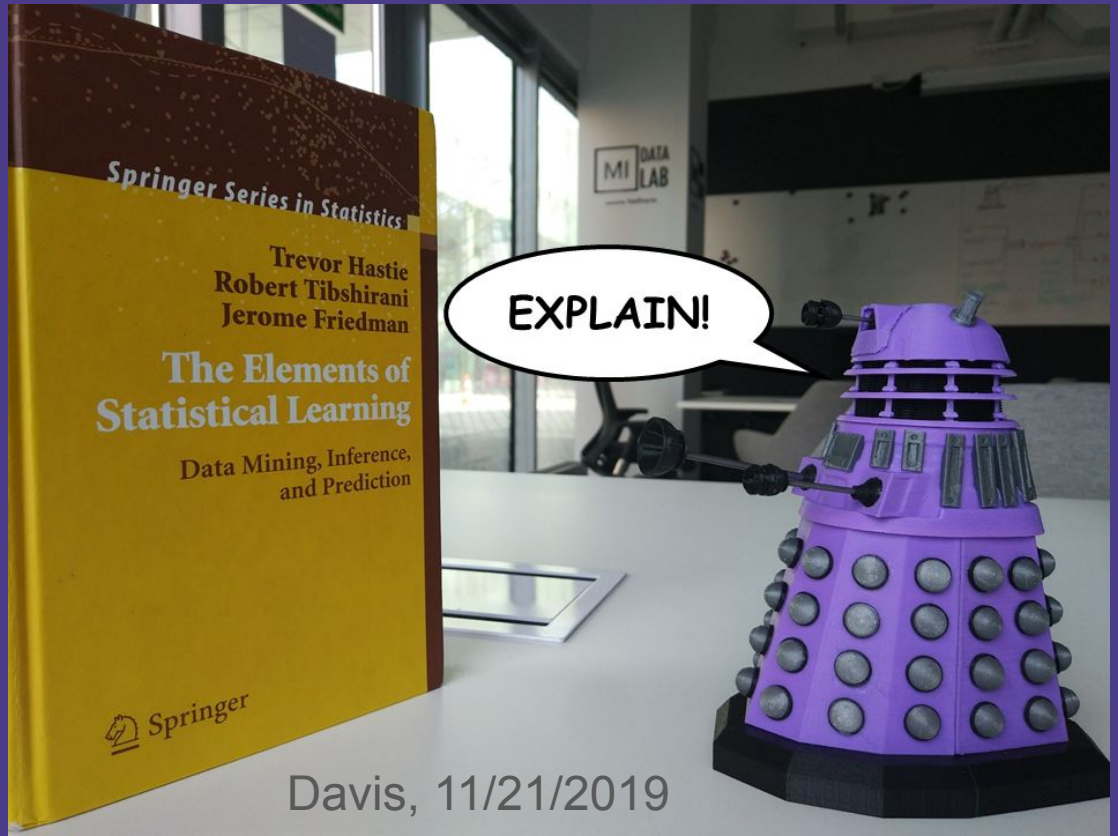
DALEX: a powerful tool for explanation of machine learning models



Davis, 11/21/2019

About me

https://github.com/agosiewska/XAI-UC-Davis

Alicja Gosiewska

PhD candidate in Interpretable Machine Learning

MSc in Mathematics (Mathematical Statistics and Data Analysis)

Faculty of Mathematics and Information Science
WARSAW UNIVERSITY OF TECHNOLOGY

MI^2 Data Lab

Warszawskie Spotkania Entuzjastów R

R-Ladies Warsaw

Intro:
Machine Learning

The success of machine learning

DeepMind's AlphaStar Final beats 99.8% of human StarCraft 2 players

KYLE WIGGERS @KYLE_L_WIGGERS OCTOBER 30, 2019 11:00 AM

Above: A screenshot of match play involving DeepMind's AlphaStar Final.

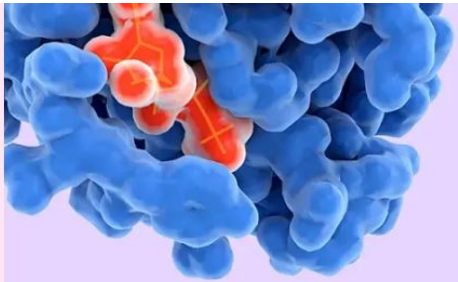
https://venturebeat.com/2019/10/30/deepminds-alphastar-final-beats-99-8-of-human-starcraft-2-players/

The success of machine learning

BLOG POST
RESEARCH

02 DEC 2018

AlphaFold: Using AI for scientific discovery



SHARE

AUTHORS

AS Andrew Senior

JJ John Jumper

Today we're excited to share DeepMind's first significant milestone in demonstrating how artificial intelligence research can drive and accelerate new scientific discoveries. With a strongly interdisciplinary approach to our work, DeepMind has brought together experts from the fields of structural biology, physics, and machine learning to apply cutting-edge techniques to predict the 3D structure of a protein based solely on its genetic sequence.

https://deepmind.com/blog/article/alphafold

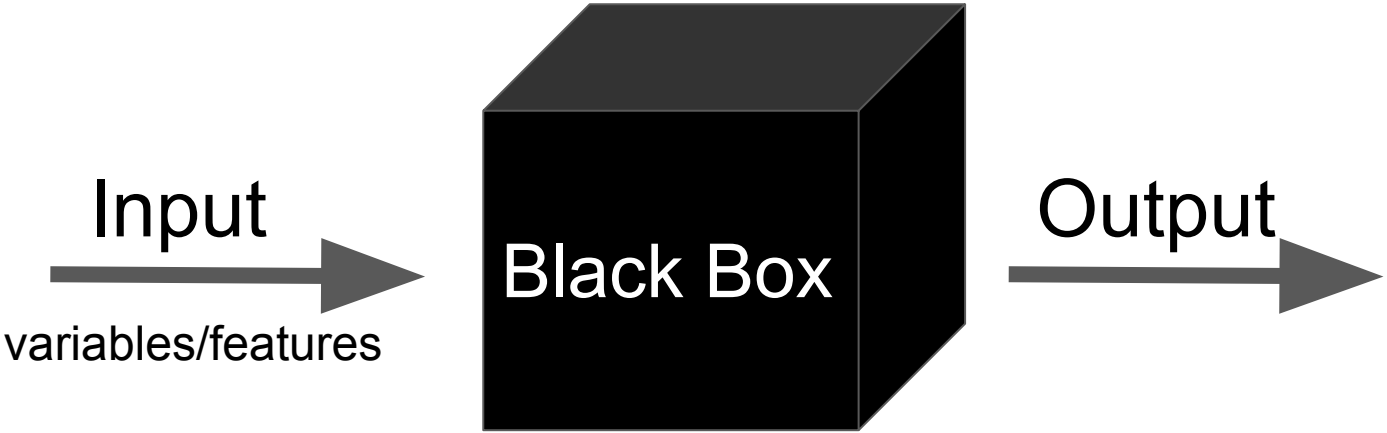
The success of machine learning

Self-Driving Cars Hit New York City Roads—But There's No Need to Panic

By Sissi Cao • 08/06/19 12:20pm

https://observer.com/2019/08/self-driving-driverless-cars-road-test-new-york-city/

Machine learning model as a black box



customer information
patient information
information about an apartment
photo
...

extend/deny a credit
start treatment or not
price of the apartment
recognition of the object on the photo
...

Problem:
Black-box Machine Learning

Apple

Apple's 'sexist' credit card investigated by US regulator

11 November 2019 Share

Steve Wozniak
@stevewoz

Replying to @dhh

The same thing happened to us. We have no separate bank accounts or credit cards or assets of any kind. We both have the same high limits on our cards, including our AmEx Centurion card. But 10x on the Apple Card.

17 6:58 AM - Nov 10, 2019

See Steve Wozniak's other Tweets

A US financial regulator has opened an investigation into claims Apple's credit card offered different credit limits for men and women.

https://www.bbc.com/news/business-50365609

COMPAS Correctional Offender Management Profiling for Alternative Sanctions

COMPAS (software)

From Wikipedia, the free encyclopedia

COMPAS, an acronym for Correctional Offender Management Profiling for Alternative Sanctions, is a case management and decision support tool developed and owned by Northpointe (now Equivant) used by U.S. courts to assess the likelihood of a defendant becoming a recidivist.[1][2]

COMPAS has been used by the U.S. states of New York, Wisconsin, California, Florida's Broward County, and other jurisdictions.[3]

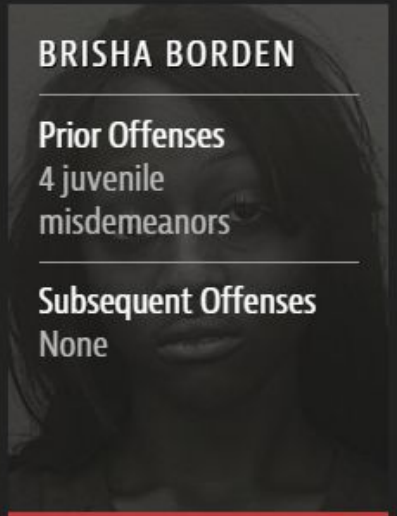
Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Two Petty Theft Arrests



Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

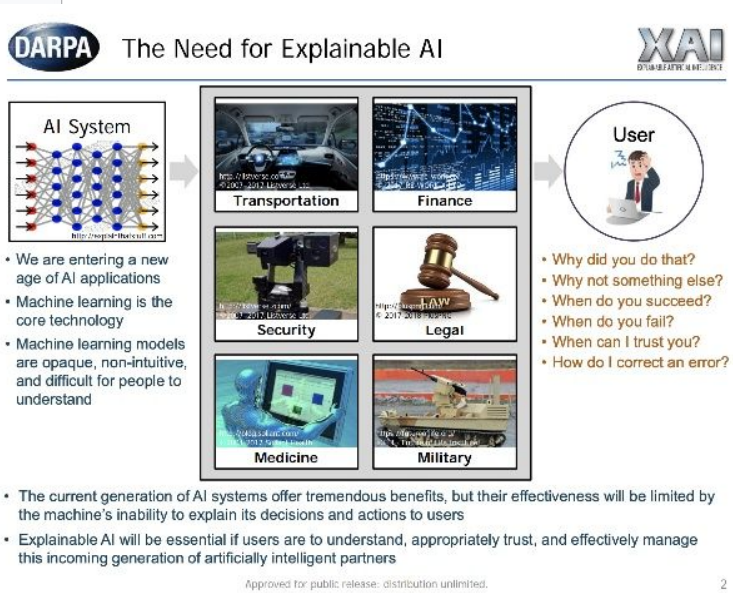
Right to explanation

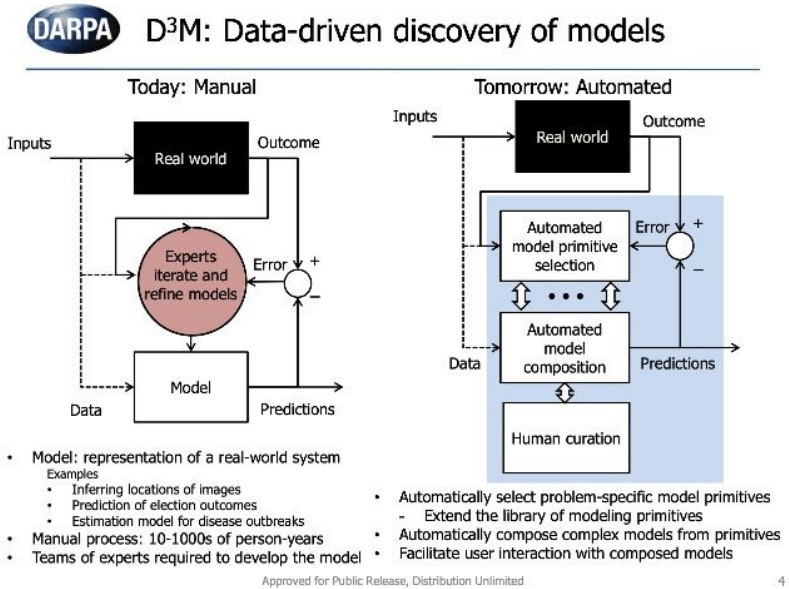
From Wikipedia, the free encyclopedia

In the regulation of algorithms, particularly artificial intelligence and its subfield of machine learning, a **right to explanation** (or **right to *an* explanation**) is a right to be given an explanation for an output of the algorithm. Such rights primarily refer to individual rights to be given an explanation for decisions that significantly affect an individual, particularly legally or financially. For example, a person who applies for a loan and is denied may ask for an explanation, which could be "Credit bureau X reports that you declared bankruptcy last year; this is the main factor in considering you too likely to default, and thus we will not give you the loan you applied for."

Some such legal rights already exist, while the scope of a general "right to explanation" is a matter of ongoing debate.

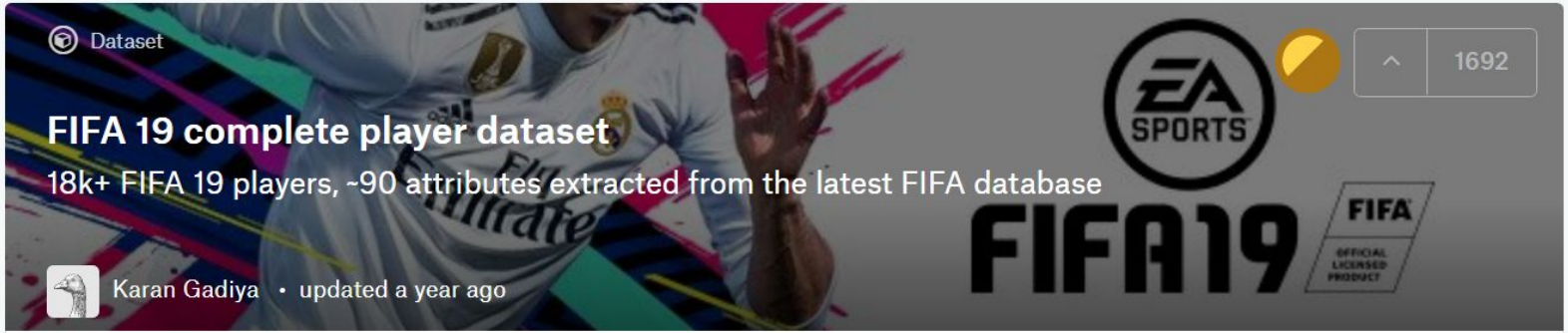
Contents [hide]





Solution:
XAI
Explainable Artificial Intelligence

kaggle Search Competitions Datasets Notebooks Discussion Courses •••



Data Kernels (308) Discussion (21) Activity Metadata Download (9 MB) New Notebook

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

Description

Context

Football analytics

https://www.kaggle.com/karangadiya/%EF%AC%81fa19/data

```
fifa_gbm <- gbm(ValueNum~.-Overall,
                data = fifa19_selected,
                n.trees = 250,
                interaction.depth = 3)
```

	Age	Overall	Special	Preferred.Foot	International.Reputation	Weak.Foot	Skill.Moves	Crossing	Finishing	HeadingAccuracy	ShortPassing	Volleys	Dribbling
L..Messi	31	94	2202	Left	5	4	4	84	95	70	90	86	97
Cristiano.Ronaldo	33	94	2228	Right	5	4	5	84	94	89	81	87	88
Neymar.Jr	26	92	2143	Right	5	5	5	79	87	62	84	84	96
De.Gea	27	91	1471	Right	4	3	1	17	13	21	50	13	18
K..De.Bruyne	27	91	2281	Right	4	5	4	93	82	55	92	82	86
E..Hazard	27	91	2142	Right	4	4	4	81	84	61	89	80	95
L..Modrić	32	91	2280	Right	4	4	4	86	72	55	93	76	90
L..Suárez	31	91	2346	Right	5	4	3	77	93	77	82	88	87
Sergio.Ramos	32	91	2201	Right	4	3	3	66	60	91	78	66	63
J..Oblak	25	90	1331	Right	3	3	1	13	11	15	29	13	12

DrWhy.AI : : grammar for model exploration

Model factory

Models are created in different languages with various libraries. New libraries will emerge, existing libraries will change.

Each library has different internal model representation, different default parameters, different format of model predictions.

Model adapter

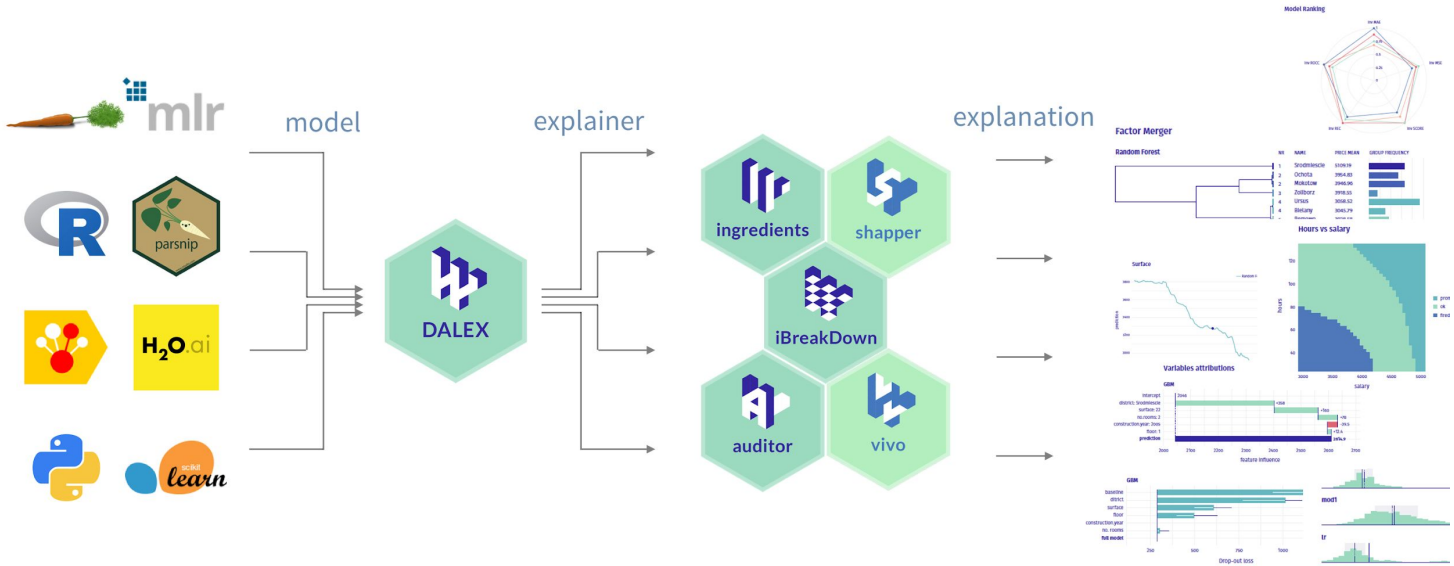
The package DALEX wraps models created by different factories into a uniform structure that can be then used by model explainers.

Function explain() is a generic interface for wrapping of models. One can use also explain_mlr() , explain_h2o(), explain_caret(), explain_scikitlearn() and other factory specific wrappers.

Model explainers

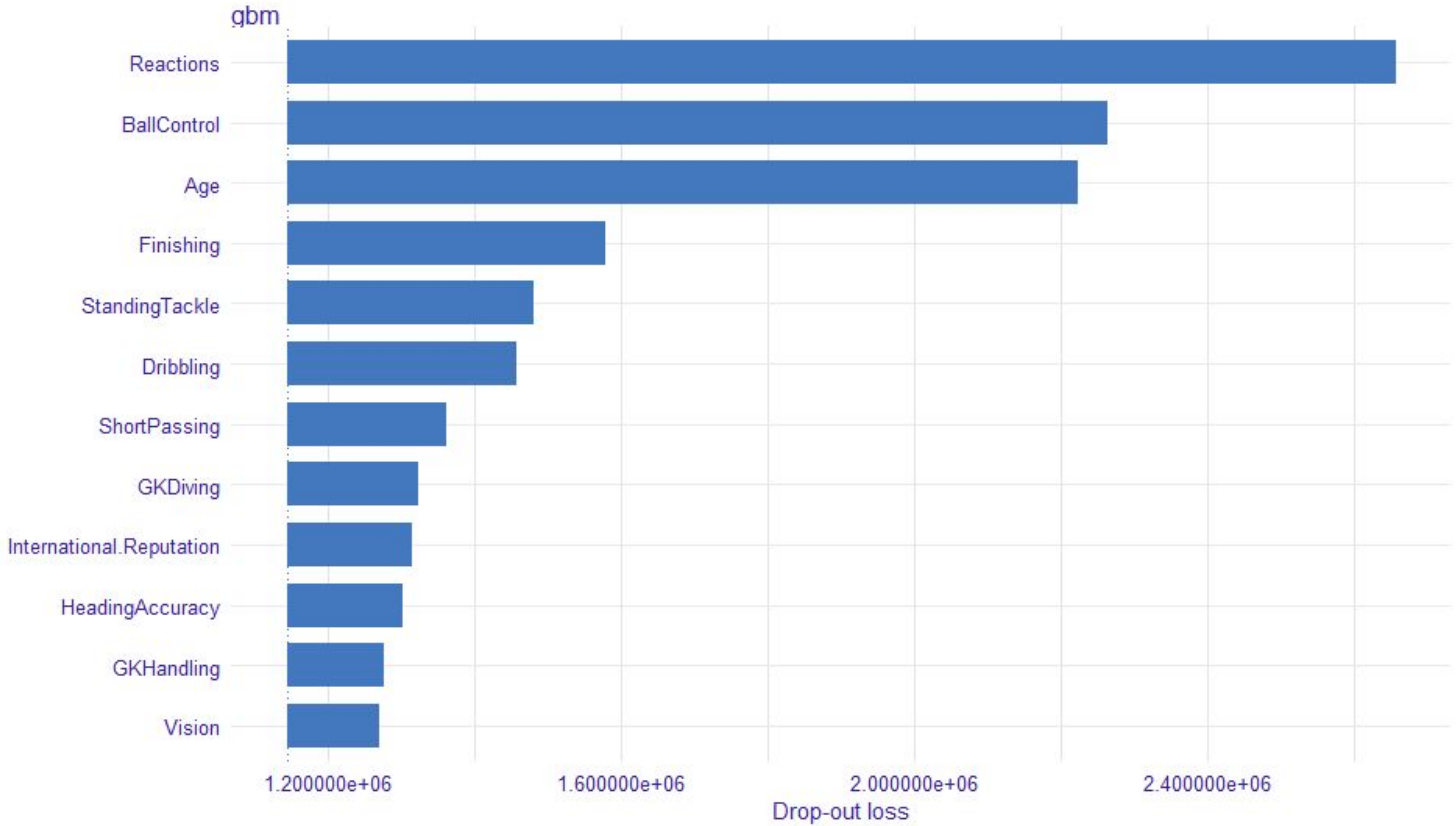
DrWhy contains collection of packages for model visual exploration, explanation and debugging. It supports for local (instance level) and global (batch level) model exploration.

Each explainer works on wrappers created with DALEX package.
For each explainer one can use generic plot function to create static chart for selected aspect of a model or generic plotD3 function for interactive version.

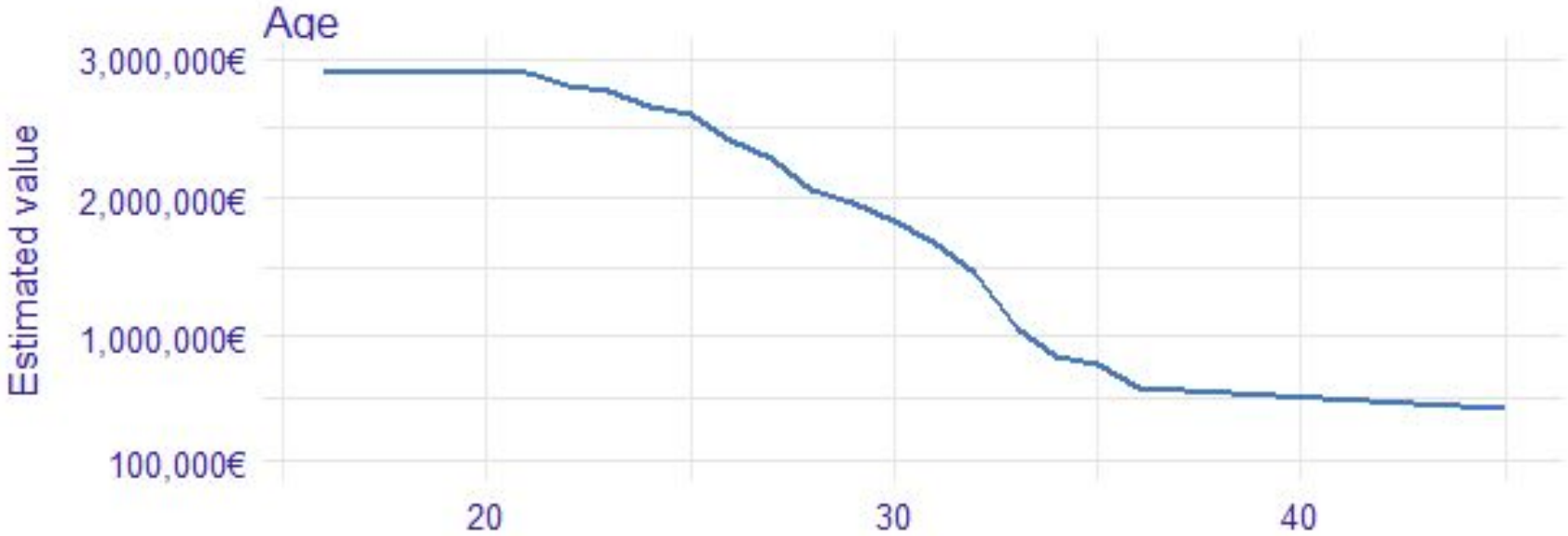


```
fifa_gbm <- gbm(ValueNum~.,
                data = fifa19_selected,
                n.trees = 250,
                interaction.depth = 3)


library("DALEX")
fifa_gbm_exp <- explain(fifa_gbm,
                        data = fifa19_selected,
                        y = fifa19_selected$ValueNum,
                        predict_function = function(m,x)
                            predict(m, x, n.trees = 250))
```

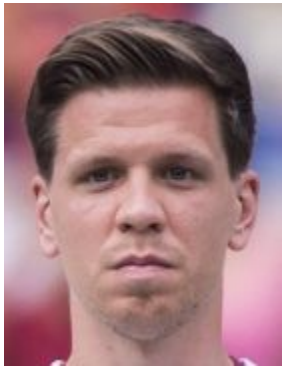
```
library("ingredients")
fifa_feat <- feature_importance(fifa_gbm_exp)
plot(fifa_feat, max_vars = 12)
```



```
fifa19_pd <- ingredients::partial_dependency(fifa_gbm_exp,
                                             variables = "Age")
plot(fifa19_pd)
```



Wojciech **Szczęsny**

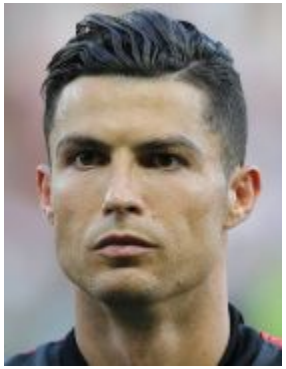


Date of birth/Age: **Apr 18, 1990 (29)**

Citizenship: **Poland**

Height: **1,96 m**

Position: **Goalkeeper**



Cristiano Ronaldo

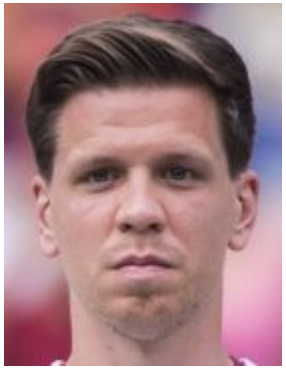
Date of birth/Age: **Feb 5, 1985 (34)**

Citizenship: **Portugal**

Height: **1,87 m**

Position: **Left Winger**

Wojciech **Szczesny**



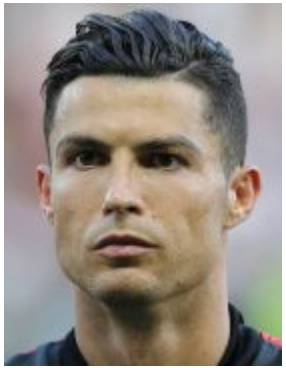
Date of birth/Age: **Apr 18, 1990 (29)**

Citizenship: **Poland**

Height: **1,96 m**

Position: **Goalkeeper**

Prediction from GBM model:
28 532 142 EUR ~ 31 605 481 USD



Cristiano Ronaldo

Date of birth/Age: **Feb 5, 1985 (34)**

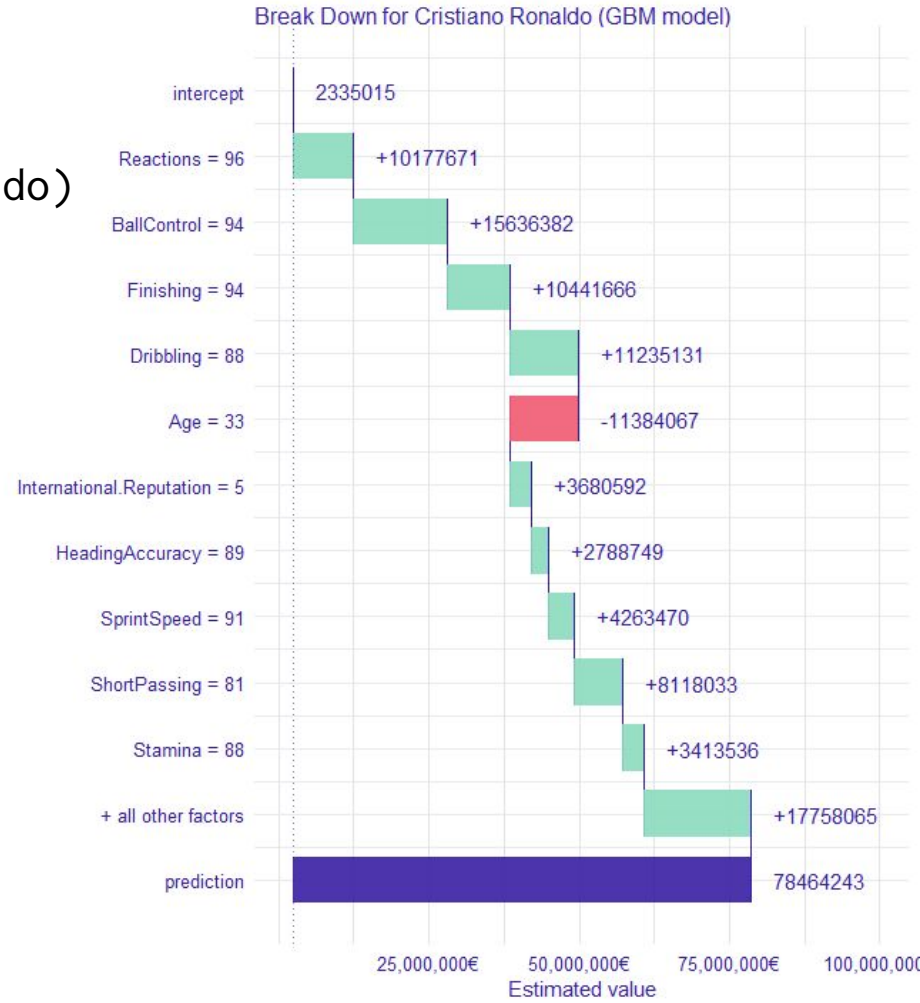
Citizenship: **Portugal**

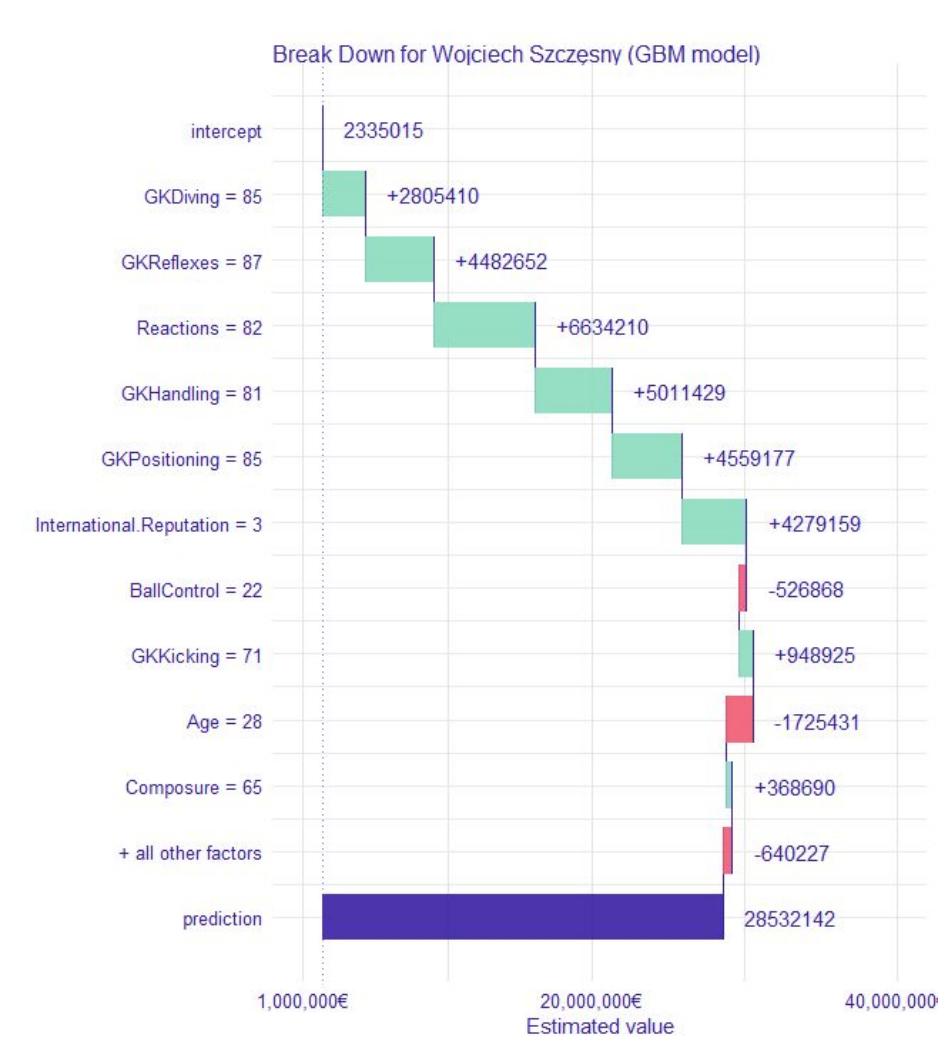
Height: **1,87 m**

Position: **Left Winger**

Prediction from GBM model:
78 464 243 EUR ~ 86 914 057 USD

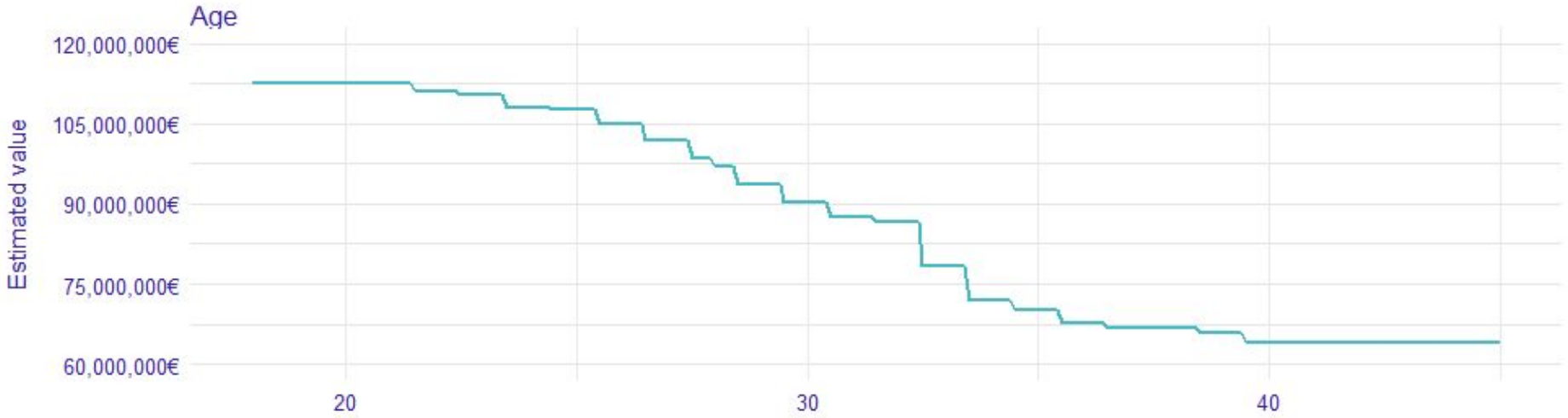
```
library("iBreakDown")

fifa_cr_gbm <- break_down(fifa_gbm_exp,
        new_observation = cristiano_ronaldo)

plot(fifa_cr_gbm)
```



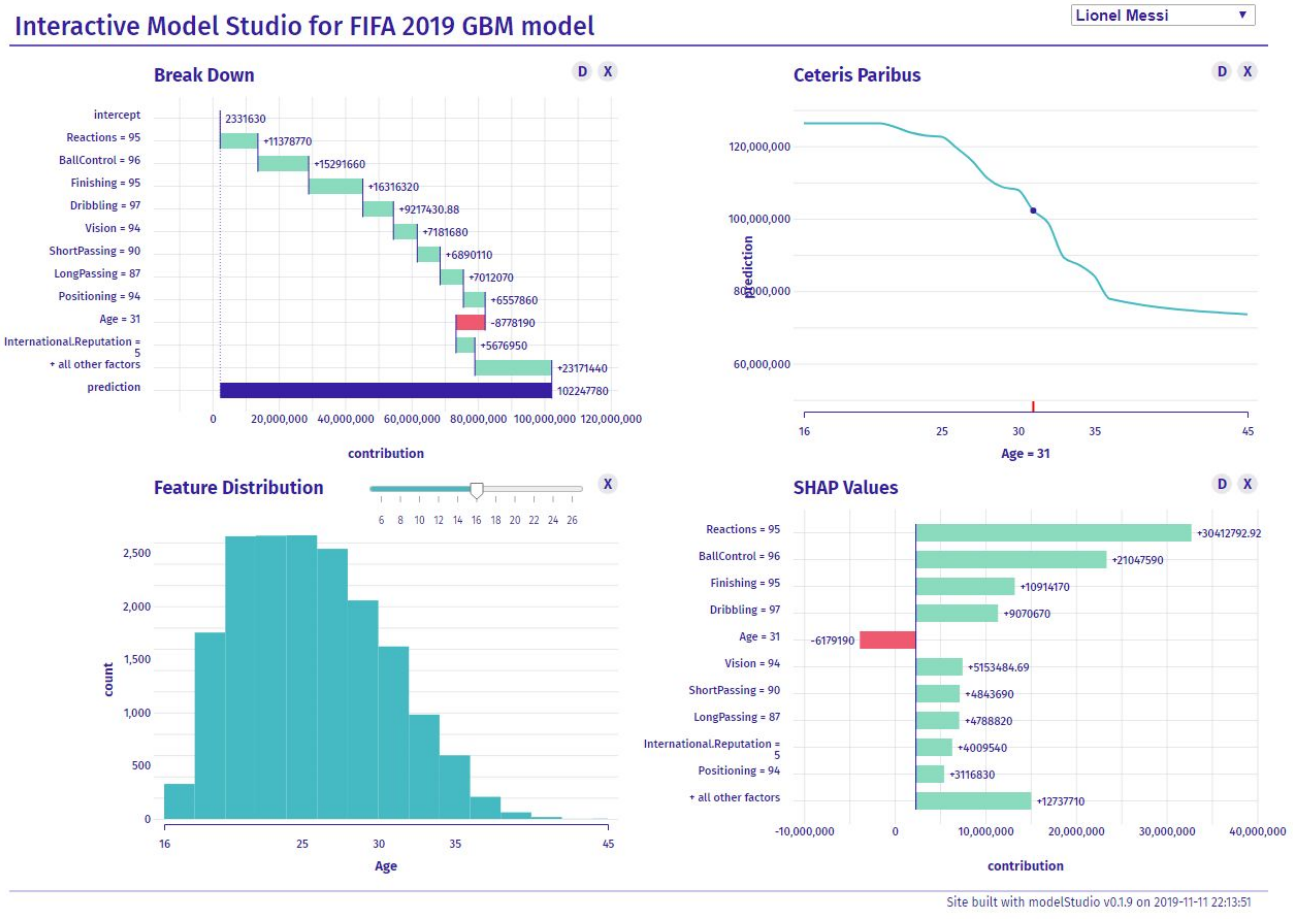

Break Down for Wojciech Szczęsny (GBM model)
intercept
2335015
GKDiving = 85
+2805410
GKReflexes = 87
+4482652
Reactions = 82
+6634210
GKHandling = 81
+5011429
GKPositioning = 85
+4559177
International.Reputation = 3
+4279159
BallControl = 22
-526868
GKKicking = 71
+948925
Age = 28
-1725431
Composure = 65
+368690
+ all other factors
-640227
prediction
28532142
1,000,000€
20,000,000€
40,000,000
Estimated value


Break Down for Cristiano Ronaldo (GBM model)
intercept
2335015
Reactions = 96
+10177671
BallControl = 94
+15636382
Finishing = 94
+10441666
Dribbling = 88
+11235131
Age = 33
-11384067
International.Reputation = 5
+3680592
HeadingAccuracy = 89
+2788749
SprintSpeed = 91
+4263470
ShortPassing = 81
+8118033
Stamina = 88
+3413536
+ all other factors
+17758065
prediction
78464243
25,000,000€
50,000,000€
75,000,000€
100,000,000
Estimated value

```
library("ingedients")
fifa19_cp_pg <- ceteris_paribus(fifa_gbm_exp,
                new_observation = cristiano_ronaldo,
                variables = "Age")
plot(fifa19_cp_pg)
```



```
modelStudio(fifa_gbm_exp,
  new_observation = rbind(messi,lewandowski,szczesny,szalai,ronaldo,neymar),
  B = 5,
  digits = 0)
```



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

More

http://drwhy.ai/

 **http://gosiewska.com/**

 **alicjagosiewska@gmail.com**

 **agosiewska**

Stay at the University of California is founded as the part of the RISE RENOIR project (call H2020-MSCA-RISE-2015, grant agreement no. 691152).