



# **1° Meetup de Data Science**

Gabriel Mariotto

cielo

## **– Agenda**

**18:00-18:15 – Recepção**

**18:15-18:30 – Abertura**

**18:30-19:15 – Interpretabilidade de Modelos Black Box (Cielo)**

**19:15-20:00 – Geração de Imagem/Áudio com GANs (neurônio.ai)**

**20:00-21:00 – Confraternização/HH**



# — Abertura

Gabriel Mariotto

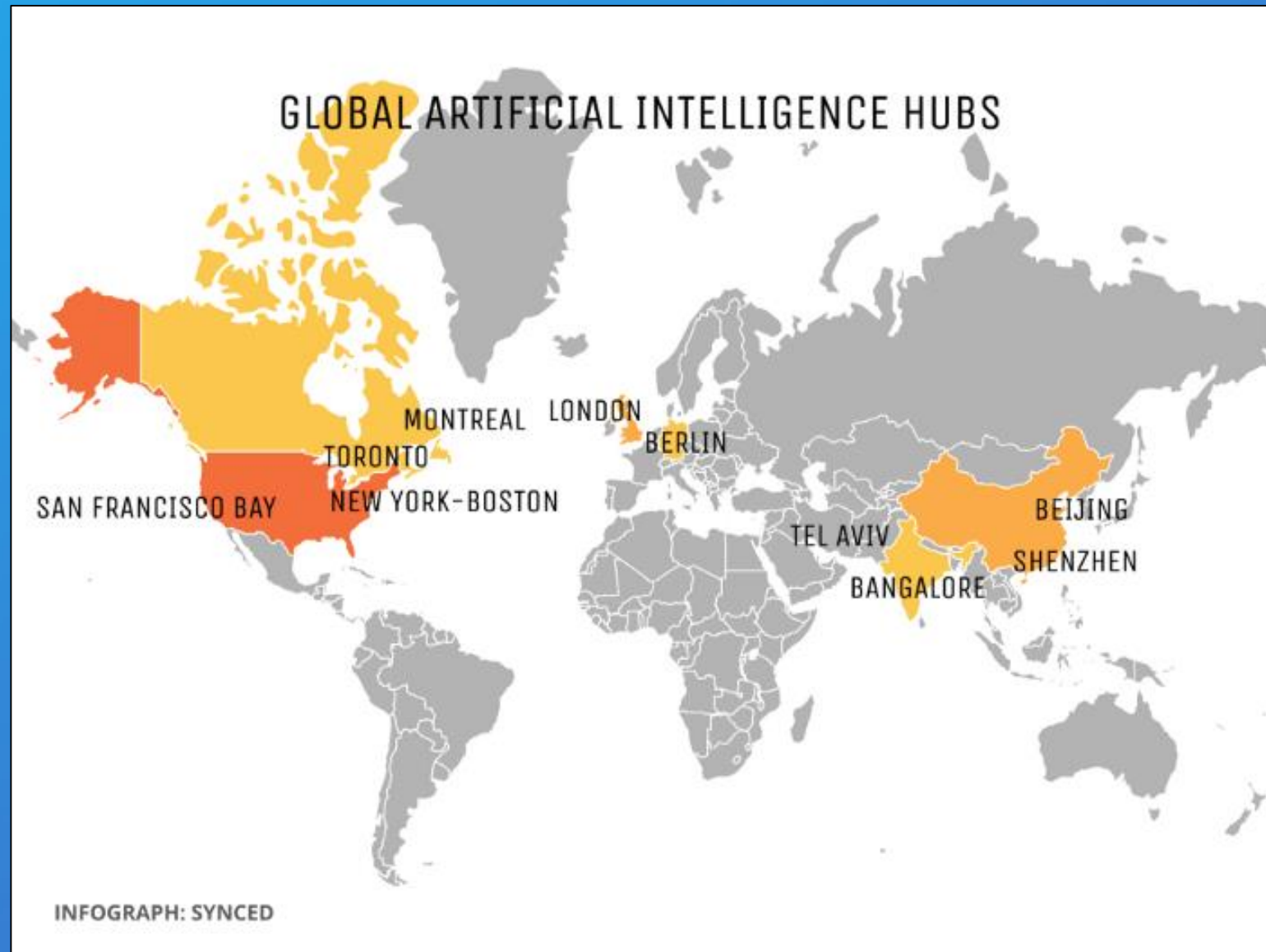
cielo



## ***– Objetivos do Meetup***

- 1. Compartilhar conhecimento**
- 2. Ideias para negócios**
- 3. HAVE FUN!**

## ***Desafio***





## GLOBAL AI CONFERENCE LIST MAY 2019–MARCH 2020

Created by Allie K. Miller

CONFERENCE NAME	CITY	COUNTRY	CONFERENCE DATES	WEBSITES
Deep Learning and Security Workshop	San Francisco	USA	May 23, 2019	<a href="https://www.ieee-security.org/1.../">https://www.ieee-security.org/1.../</a>
Malta AI & Blockchain Summit	Misida	Malta	May 23 - 24, 2019	<a href="https://malta.blockchainsummit.com/">https://malta.blockchainsummit.com/</a>
Rework Deep Learning in Healthcare Summit	Boston	USA	May 23 - 24, 2019	<a href="https://www.re-work.co/events/.../">https://www.re-work.co/events/.../</a>
NAACL-HLT	Minneapolis	USA	June 2 - 7, 2019	<a href="https://naacl2019.org/">https://naacl2019.org/</a>
NDR Artificial Intelligence Conference	Bucharest	Romania	June 4 - 6, 2019	<a href="https://ndrconf.ai/">https://ndrconf.ai/</a>
re:MARS by Amazon	Las Vegas	USA	June 4 - 7, 2019	<a href="https://remars.amazon.com/">https://remars.amazon.com/</a>
Big Data & Analytics for Retail Summit	Chicago	USA	June 7 - 8, 2019	<a href="https://www.theinnovationenterprise.com/.../">https://www.theinnovationenterprise.com/.../</a>
International Conference on Machine Learning (ICML)	Long Beach	USA	June 10 - 15, 2019	<a href="https://icml.cc/Conferences/2019">https://icml.cc/Conferences/2019</a>
AI Summit	London	United Kingdom	June 11 - 13, 2019	<a href="https://london.theaisummit.com/">https://london.theaisummit.com/</a>
Microsoft Azure + AI Conference	Orlando	USA	June 11 - 13, 2019	<a href="https://azureaiconf.com/#/">https://azureaiconf.com/#/</a>
AI Toronto	Toronto	Canada	June 12 - 13, 2019	<a href="https://ai-toronto.com/2019/">https://ai-toronto.com/2019/</a>
Data Science Salon (Media and Entertainment)	New York	USA	June 13, 2019	<a href="https://datascience.salon/newyork/">https://datascience.salon/newyork/</a>
Deep Learning World	Las Vegas	USA	June 16 - 20, 2019	<a href="https://www.deeplearningworld.com/">https://www.deeplearningworld.com/</a>
Computer Vision and Pattern Recognition (CVPR)	Long Beach	USA	June 16 - 21, 2019	<a href="http://cvpr2019.thecvf.com/">http://cvpr2019.thecvf.com/</a>
O'Reilly Artificial Intelligence Conference	Beijing	China	June 18 - 21, 2019	<a href="https://conferences.oreilly.com/.../">https://conferences.oreilly.com/.../</a>
AI & Big Data Expo Europe	Amsterdam	Netherlands	June 19 - 20, 2019	<a href="https://www.ai-expo.net/europe/">https://www.ai-expo.net/europe/</a>
AI Dev Summit at Developer Week	New York	USA	June 19 - 20, 2019	<a href="https://www.developerweek.com/NYC/">https://www.developerweek.com/NYC/</a>
E66 NYC	New York	USA	June 20, 2019	<a href="https://nyc.eggdataib.com/">https://nyc.eggdataib.com/</a>
RE.WORK AI for Good Summit	San Francisco	USA	June 20 - 21, 2019	<a href="https://www.re-work.co/events/.../">https://www.re-work.co/events/.../</a>
Rework Applied AI Summit	San Francisco	USA	June 20 - 21, 2019	<a href="https://www.re-work.co/events/.../">https://www.re-work.co/events/.../</a>
Ai4 Retail	New York	USA	June 24 - 25, 2019	<a href="https://ai4io/retail/">https://ai4io/retail/</a>
AI World Government	Washington, DC	USA	June 24 - 26, 2019	<a href="https://www.aiworldgov.com/">https://www.aiworldgov.com/</a>
Conference on Learning Theory (COLT)	Phoenix	USA	June 25 - 28, 2019	<a href="http://learningtheory.org/colt2019/">http://learningtheory.org/colt2019/</a>
AIIME Conference on Artificial Intelligence in Medicine	Poznan	Poland	June 26 - 29, 2019	<a href="http://aime19.aimedicine.info/">http://aime19.aimedicine.info/</a>
Interpol World	Singapore	Singapore	July 2 - 4, 2019	<a href="https://www.interpol-world.com/">https://www.interpol-world.com/</a>

# Desafio

## GLOBAL AI CONFERENCE LIST MAY 2019–MARCH 2020

Created by Allie K. Miller

CONFERENCE NAME	CITY	COUNTRY	CONFERENCE DATES	WEBSITES
Deep Learning and Security Workshop	San Francisco	USA	May 21, 2019	<a href="#">https://www.usenix.org/conference/sec19</a>
Meta AI: Blockchain Summit	Malta	Malta	May 21–23, 2019	<a href="#">https://metaai.io/blockchain-summit</a>
Machine Learning in the Media Summit	Paris	USA	May 21–23, 2019	<a href="#">https://www.mlmachinelearning.com</a>
AMICI ML7	Minneapolis	USA	June 1–7, 2019	<a href="#">https://amici.ai</a>
MOA AI/ML Intelligence Conference	Buenos Aires	Argentina	June 4–8, 2019	<a href="#">https://moa.ai</a>
IoT ML by Design	Los Angeles	USA	June 4–7, 2019	<a href="#">https://www.iotmlbydesign.com</a>
Big Data & Analytics for Retail Summit	Chicago	USA	June 7–8, 2019	<a href="#">https://www.bigdataanalyticsforretail.com</a>
International Conference on Machine Learning (ICML)	Long Beach	USA	June 9–15, 2019	<a href="#">https://icml.cc/2019</a>
AI Summit	London	United Kingdom	June 11–13, 2019	<a href="#">https://aismart.co.uk</a>
MetaHack AI/ML @ San Francisco	Ottawa	USA	June 11–13, 2019	<a href="#">https://metahack.ai</a>
AI Talent	Toronto	Canada	June 11–13, 2019	<a href="#">https://aitalent.ca</a>
Data Science Talent (Data & DataScience)	New York	USA	June 12, 2019	<a href="#">https://www.dataanddatascience.com</a>
Deep Learning (DL)	Los Angeles	USA	June 12–13, 2019	<a href="#">https://www.deeplearning.com</a>
Computer Vision and Pattern Recognition (CVPR)	Long Beach	USA	June 16–21, 2019	<a href="#">https://cvpr.thecvf.com</a>
2019 AI/ML Intelligence Conference	Bellingham	China	June 18–21, 2019	<a href="#">https://www.aimeconline.com</a>
AI & Big Data Expo Europe	Amsterdam	Netherlands	June 18–19, 2019	<a href="#">https://www.aibigdataexpo.com</a>
AI Data Science in Europe: Week	New York	USA	June 19–24, 2019	<a href="#">https://www.aidata-science.com</a>
USO AI/ML	New York	USA	June 24, 2019	<a href="#">https://uso.ai</a>
AI/ML Summit @ San Francisco	San Francisco	USA	June 24–25, 2019	<a href="#">https://www.aismart.co.uk</a>
Future Applied AI Summit	San Francisco	USA	June 24–25, 2019	<a href="#">https://www.futureappliedai.com</a>
AI Summit	New York	USA	June 24–25, 2019	<a href="#">https://ai-summit.com</a>
AI World Development	Washington, DC	USA	June 24–25, 2019	<a href="#">https://aiworlddev.com</a>
Conference on Learning Theory (COLT)	Paris	France	June 24–25, 2019	<a href="#">https://colt.acm.org</a>
AI/ML Conference on Artificial Intelligence in Medicine	Palma	Spain	June 24–25, 2019	<a href="#">https://aiim2019.com</a>
Intelligence World	Singapore	Singapore	July 1–4, 2019	<a href="#">https://intelligenceworld.com</a>
TensorFlow ML	San Francisco	USA	July 16–17, 2019	<a href="#">https://www.tensorflowml.com</a>
2019 International Summit on ML	New York	USA	July 16–18, 2019	<a href="#">https://www.2019summitonml.com</a>
Machine Learning Intelligence Conference on ML/ML	Chengdu	China	July 16–18, 2019	<a href="#">https://www.mlintelligence.com</a>
Data Science Summit	Singapore	Singapore	July 17–18, 2019	<a href="#">https://www.data-science-summit.com</a>
ML/ML International Conference on Machine Learning and Data Mining	New York	USA	July 24–25, 2019	<a href="#">https://www.mlml.com</a>
Summit on Computer and Language (SCOL)	Paris	France	July 24–25, 2019	<a href="#">https://www.scol.com</a>
Intelligence Applied Cognitive Intelligence (IACI)	Seattle	USA	August 1–4, 2019	<a href="#">https://www.iaci.com</a>
AI/ML Summit @ San Francisco	San Francisco	USA	August 18–20, 2019	<a href="#">https://www.aismart.co.uk</a>
Global AI Data Conference	San Jose	USA	August 18–20, 2019	<a href="#">https://www.globalai-data.com</a>
AI Summit	New York	USA	August 21–22, 2019	<a href="#">https://ai-summit.com</a>
Future Applied AI Summit	New York	USA	September 4–8, 2019	<a href="#">https://www.futureappliedai.com</a>
Black Machine Data Conference	San Jose	USA	September 9–10, 2019	<a href="#">https://www.blackmachine.com</a>
AI/ML	London	United Kingdom	September 9–10, 2019	<a href="#">https://ai-ml.com</a>
Data Science Talent (Data & DataScience)	London	USA	September 9–10, 2019	<a href="#">https://www.dataanddatascience.com</a>
AI Summit Singapore	Singapore	Singapore	September 10–11, 2019	<a href="#">https://ai-summit.com</a>
USO AI/ML	New York	USA	September 10–11, 2019	<a href="#">https://uso.ai</a>
The 2019 Conference on Artificial Intelligence Systems	San Jose	USA	September 10–11, 2019	<a href="#">https://www.2019conference.com</a>
Future Applied AI Summit	London	United Kingdom	September 10–11, 2019	<a href="#">https://www.futureappliedai.com</a>
2019 International Conference on Data Mining (ICDM)	Taipei	Taiwan	September 12–15, 2019	<a href="#">https://icdm.acm.org</a>
7th Global Summit on AI/ML Intelligence & Neural Networks	Dubai	UAE	September 22–24, 2019	<a href="#">https://www.7thglobalsummit.com</a>
Data Science Summit	Buenos Aires	Argentina	September 23–24, 2019	<a href="#">https://www.data-science-summit.com</a>
Global AI/ML Intelligence Conference	Paris	France	October 1–3, 2019	<a href="#">https://www.globalai-ml.com</a>
World Summit AI	Amsterdam	Netherlands	October 8–10, 2019	<a href="#">https://www.worldsummitai.com</a>
2019 ML Workshop on Machine Learning for Signal Processing	Philadelphia	USA	October 12–13, 2019	<a href="#">https://mlworkshop.com</a>
2019 AI Conference	London	United Kingdom	October 14–15, 2019	<a href="#">https://ai-conference.com</a>
Data Science Talent (Data & DataScience) P&G	Seattle	USA	October 14–15, 2019	<a href="#">https://www.dataanddatascience.com</a>
AI/ML Summit @ San Francisco	New York	USA	October 15, 2019	<a href="#">https://www.aismart.co.uk</a>
Future Applied AI Summit	New York	USA	October 15–16, 2019	<a href="#">https://www.futureappliedai.com</a>
World Machine Learning and Deep Learning Congress	London	United Kingdom	October 15–16, 2019	<a href="#">https://www.worldmachinelearning.com</a>
International Conference on Computer Vision (ICCV)	Seoul	South Korea	October 27–November 1, 2019	<a href="#">https://iccv.thecvf.com</a>
Deep Learning in the Real World: Planning (DLRP)	Long Beach	USA	November 3–7, 2019	<a href="#">https://www.dlrp.com</a>
International Conference on Intelligence, Robotics and Systems (IIRS)	Wuhan	China	November 3–7, 2019	<a href="#">https://www.iirs.com</a>
Data Science Talent (Data & DataScience) P&G	Los Angeles	USA	November 4, 2019	<a href="#">https://www.dataanddatascience.com</a>
AI Summit	New York	USA	November 11–12, 2019	<a href="#">https://ai-summit.com</a>
Singapore's Future Summit	Singapore	Singapore	November 11–12, 2019	<a href="#">https://www.singaporefuture.com</a>
Deep Learning Summit	Washington	USA	November 11–12, 2019	<a href="#">https://www.deeplearning.com</a>
AI & Big Data Expo	New York	USA	November 11–12, 2019	<a href="#">https://www.aibigdataexpo.com</a>
Data Science Summit @ San Francisco	New York	USA	December 1–5, 2019	<a href="#">https://www.data-science-summit.com</a>
New AI/ML Intelligence Systems (New AI/ML)	New York	USA	December 4–5, 2019	<a href="#">https://www.newai-ml.com</a>
ML Summit @ San Francisco	New York	USA	January 1–14, 2020	<a href="#">https://www.mlsummit.com</a>
Future Applied AI Summit	San Francisco	USA	January 28–30, 2020	<a href="#">https://www.futureappliedai.com</a>
Machine Learning Summit	San Francisco	USA	January 28–30, 2020	<a href="#">https://www.machinelearning.com</a>
AI & Big Data Expo Global	London	United Kingdom	March 11–13, 2020	<a href="#">https://www.aibigdataexpo.com</a>
Future Applied AI Summit	London	United Kingdom	March 11–13, 2020	<a href="#">https://www.futureappliedai.com</a>

## Eventos por Região

- North America 21
- Europe 11
- Asia 6
- Middle East 1
- South America 1

# Cielo Big Data

A visão de Big Data da Cielo foi estruturada para alavancar uma base completa de informações sobre o varejo

Exemplos de informações capturadas



Presença em  
99,9% dos  
municípios  
brasileiros



Atuação nos principais  
setores da economia



10% do PIB e 14% do  
Consumo das Famílias



6,2 bilhões  
de vendas



Mais de 1,2 milhão de  
pontos de venda ativos



## **– Agenda**

**18:00-18:15 – Recepção**

**18:15-18:30 – Abertura**

**18:30-19:15 – Interpretabilidade de Modelos Black Box (Cielo)**

**19:15-20:00 – Geração de Imagem/Áudio com GANs (neurônio.ai)**

**20:00-21:00 – Confraternização/HH**



# **— Interpretabilidade** de modelos Black Box

1º Meetup de Data Science – 17/07/19

cielo

# — Quem Somos



**Igor Gushiken**  
**Gerente de Data Science**

- *Eng. da Computação (POLI-USP)*
- *MS Engineering (École Centrale Nantes)*

*Igor trabalha na área de dados da Cielo há 5 anos. Atualmente, seu time é responsável pelas iniciativas na Cielo envolvendo modelagem preditiva e prescritiva, apoiando todas as áreas da empresa.*



**Victor Fernandes**  
**Data Scientist**

- *Eng. Mecânico-Aeronáutico (ITA)*

*Victor iniciou sua carreira como Data Scientist há 2 anos, em um grande banco brasileiro. Atualmente, é Data Scientist na Cielo, onde trabalha com modelos de crédito, scoragem de leads e text mining.*

# \_O que fazemos?

## **Apoio Interno**

Suporte a  
outras áreas  
da Cielo

## **Disseminação**

Tornar a  
empresa  
"Data-Driven"

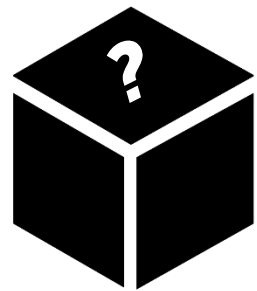
## **Machine Learning**

Modelos  
preditivos e  
prescritivos

## **Monetização**

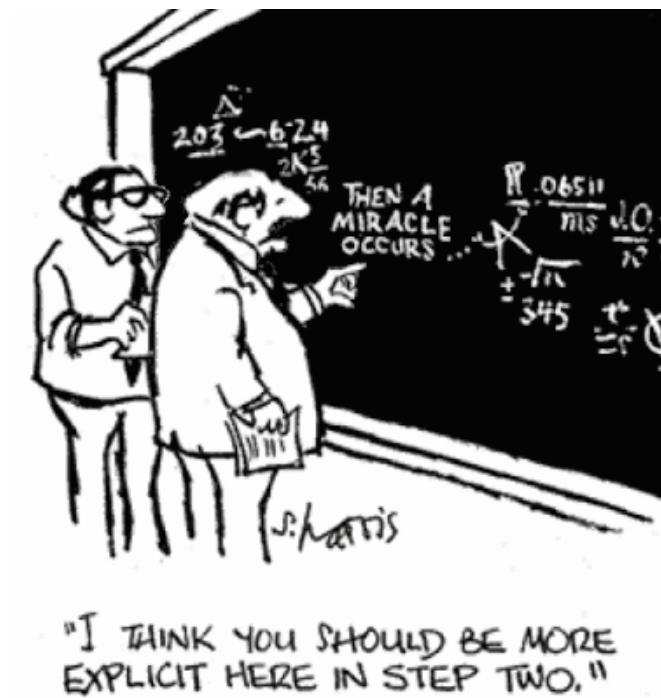
Produtos de  
dados de  
mercado

# **Interpretabilidade** de modelos Black Box





# Interpretabilidade



**Interpretabilidade** é o grau com qual um **humano** consegue entender a **razão de uma decisão**<sup>1</sup>

1) Miller, Tim. "Explanation in artificial intelligence: Insights from the social sciences." arXiv Preprint arXiv:1706.07269. (2017).

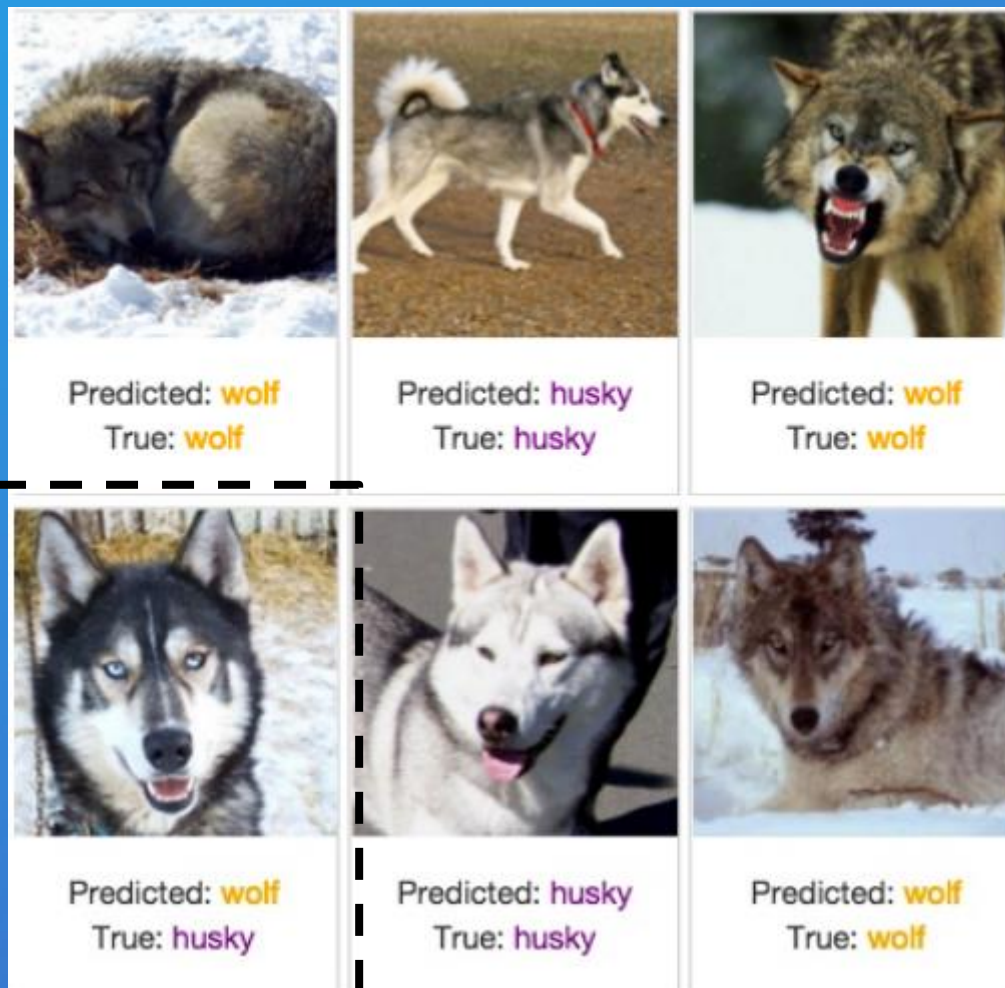
# ***Interpretabilidade de Modelos***

***Por que devemos interpretar modelos de Machine Learning?***

1. *Garantir confiança no modelo*
2. *Explicar decisões e gerar insights*
3. *Satisfazer requisitos regulatórios (LGPD/GDPR)*
4. *Debugar/melhorar o modelo*

# Acurácia é suficiente para garantir confiança?

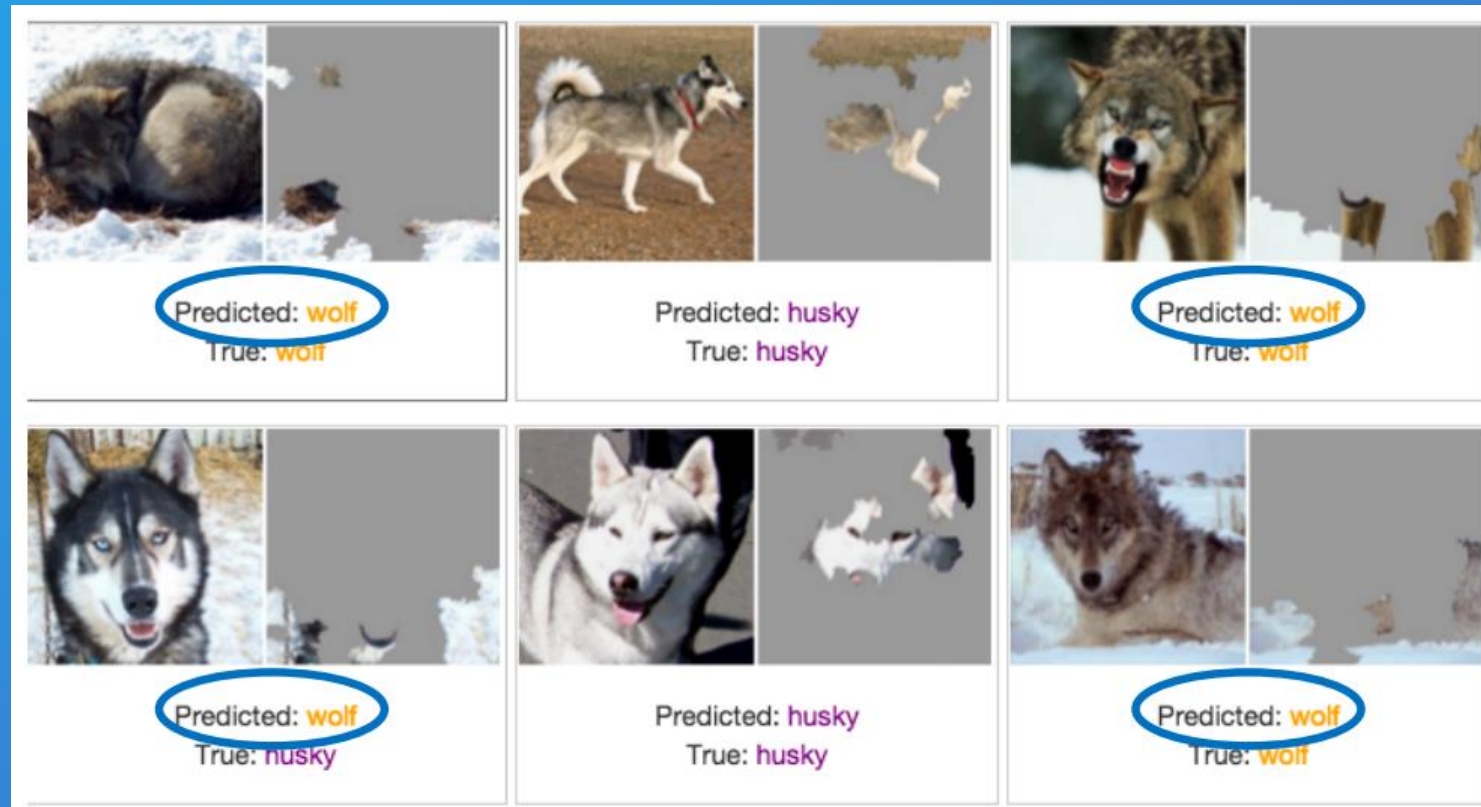
Exemplo de detector de Husky ou Lobos (artigo do LIME)



Apenas  
1 erro

# Acurácia é suficiente para garantir confiança?

Exemplo de detector de Husky ou Lobos (artigo do LIME)



Criou-se um excelente **detector de neve!**

# ***\_ Tipos de Interpretabilidade de Modelos***

## **1. Global**

*Como o modelo faz previsões a partir das variáveis de entrada?*

*Quais características de um cliente fazem com ele tenha maior propensão a churn?*

## **Interpretabilidade**

## **2. Local**

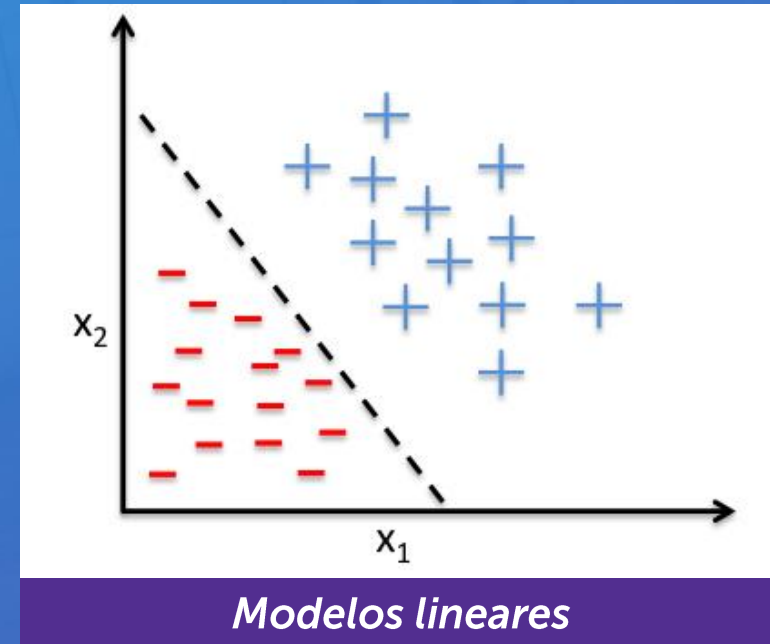
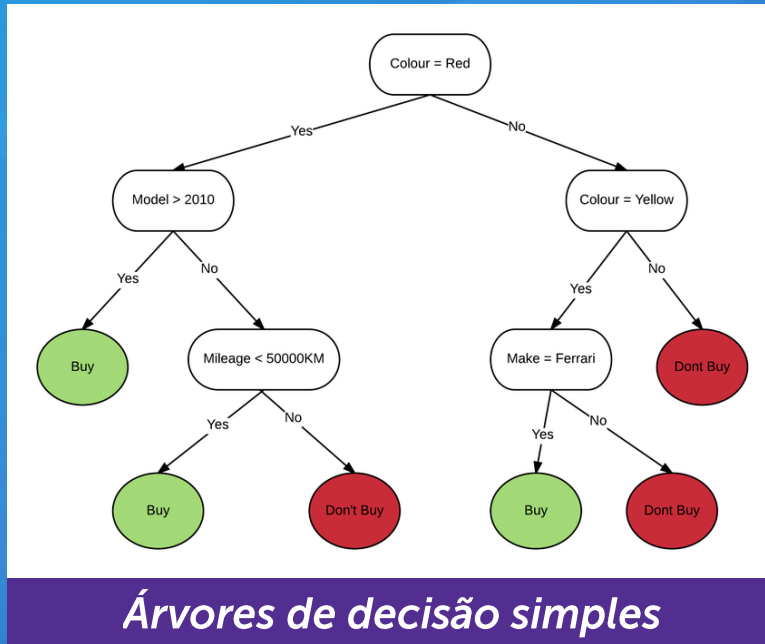
*Por que o modelo tomou essa decisão para esse caso?*

*Por que o modelo indicou que esse cliente tem alta propensão a churn?*

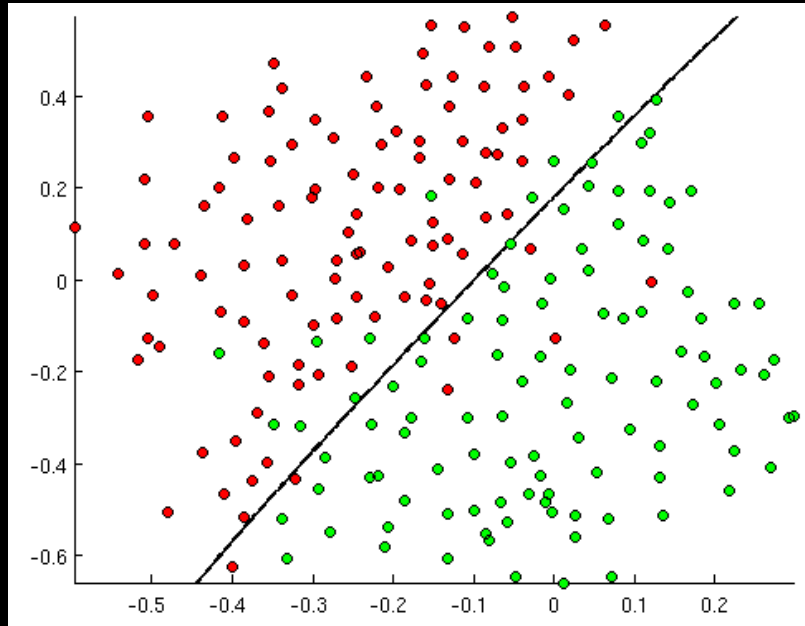


# ***Como garantir interpretabilidade de modelos?***

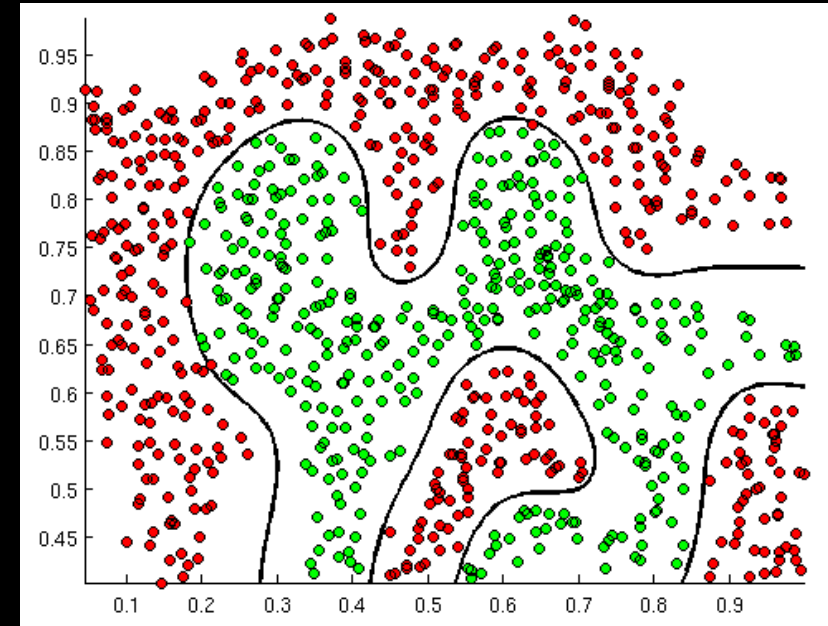
## ***1. Interpretabilidade intríntrica (modelos simples)***



**— O mundo real ~~não é um Toddynho gelado~~ é COMPLEXO!**

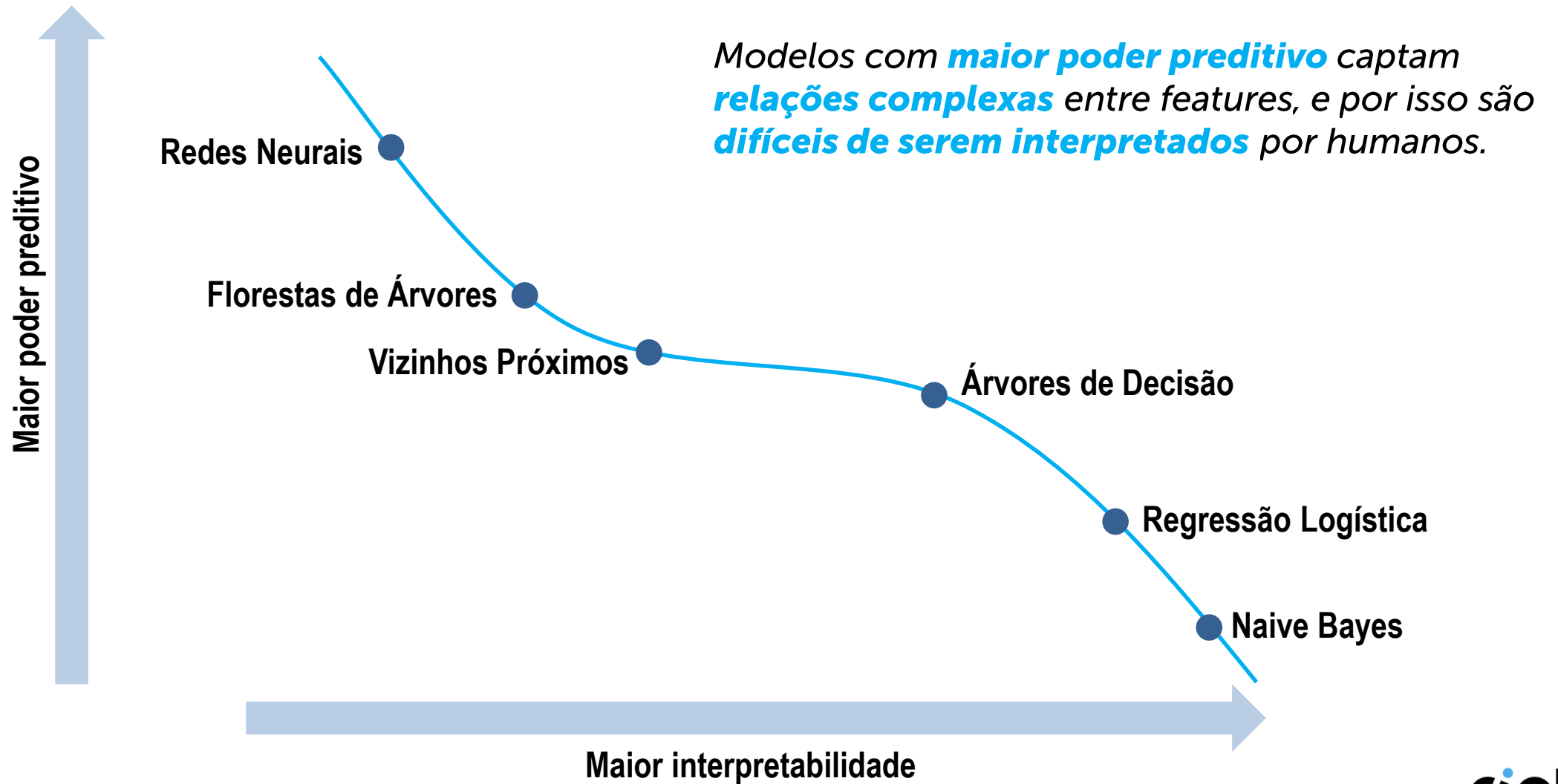


**EXPECTATIVA**

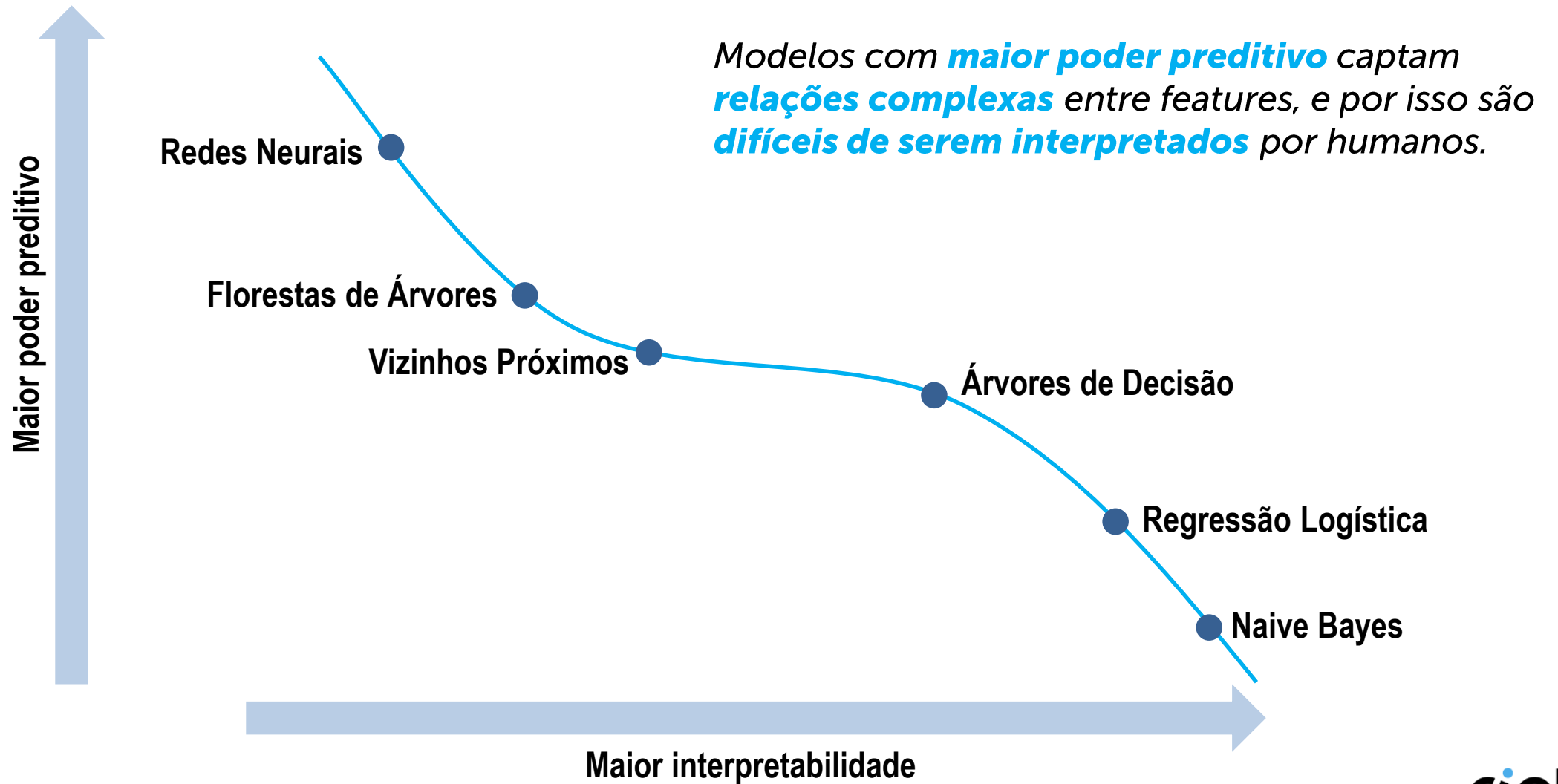


**REALIDADE**

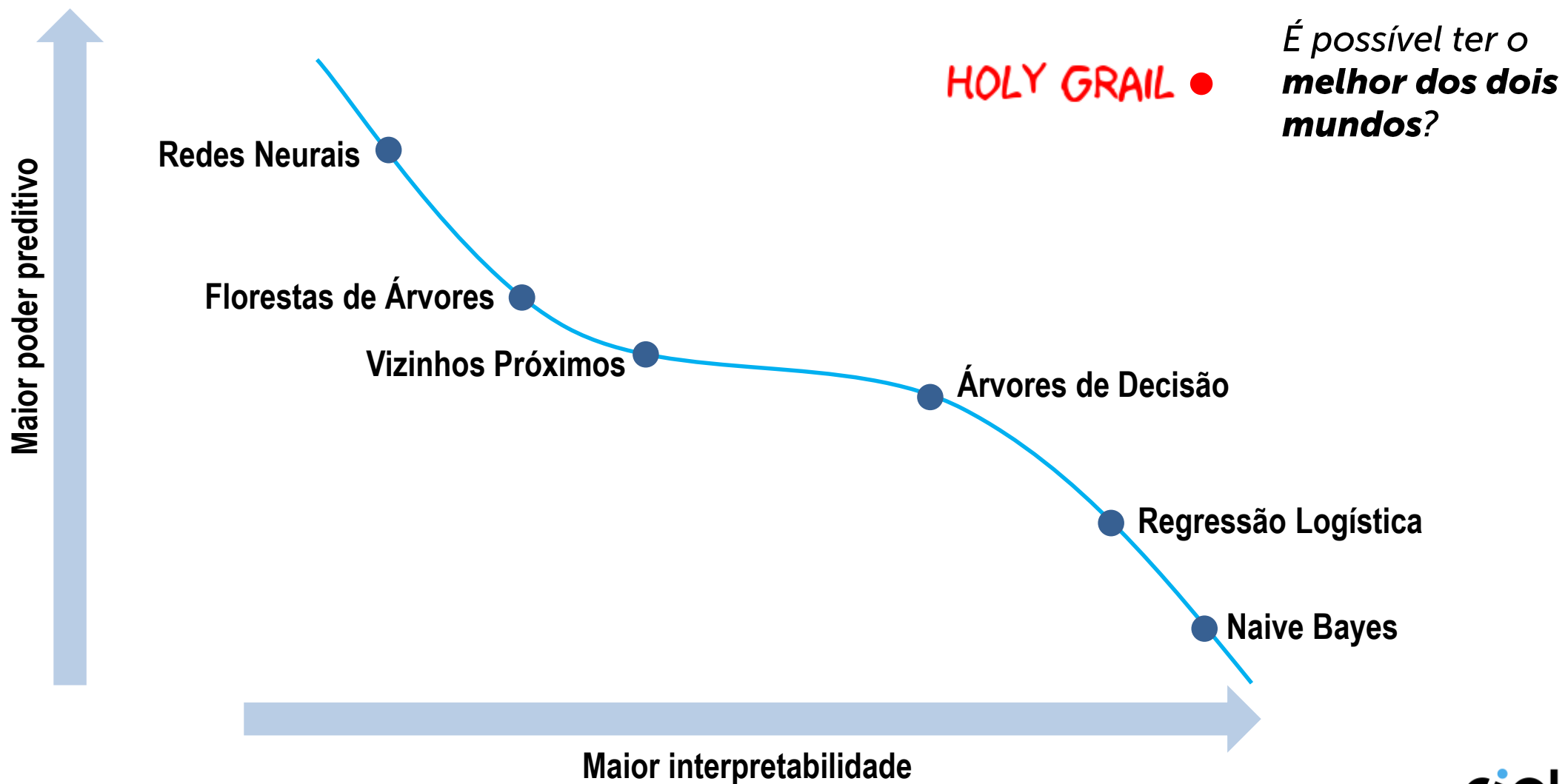
# ***Trade-off acurácia vs interpretabilidade***



# ***Trade-off acurácia vs interpretabilidade***



# ***Trade-off acurácia vs interpretabilidade***





# ***Técnicas de interpretabilidade***

## **1. Técnicas de análises exploratórias**

- *Dependências parciais*
- *Individual Conditional Expectation (ICE)*
- *Global Surrogate Models*

## **2. Pacotes de interpretabilidade de modelos (Explainers)**

- *LIME*
- *SHAP*
- *eli5*



**CODE DEMO**

# — Técnicas de interpretabilidade

## Base e modelo utilizados na demonstração



### Base de Dados Salariais: US Census Bureau – 1994

Ex: Idade, Nacionalidade, Nível de Educação, Ocupação, etc  
32,5 mil registros



### Modelo Preditivo

Indivíduo ganha mais do que \$ 50k anuais?  
Ensemble model utilizando lightgbm

	Age	Workclass	fnlwgt	Education	Education-Num	Marital Status	Occupation	Relationship	Capital Gain	Capital Loss	Hours per week	Country
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	2174	0	40	United-States
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	0	0	13	United-States
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	0	0	40	United-States
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	0	0	40	United-States

# ***Técnicas de interpretabilidade***

## **1. Técnicas de análises exploratórias**

- *Dependências parciais*
- *Individual Conditional Expectation (ICE)*
- *Global Surrogate Models*

## **2. Pacotes de interpretabilidade de modelos (Explainers)**

- *LIME*
- *SHAP*
- *eli5*



**CODE DEMO**

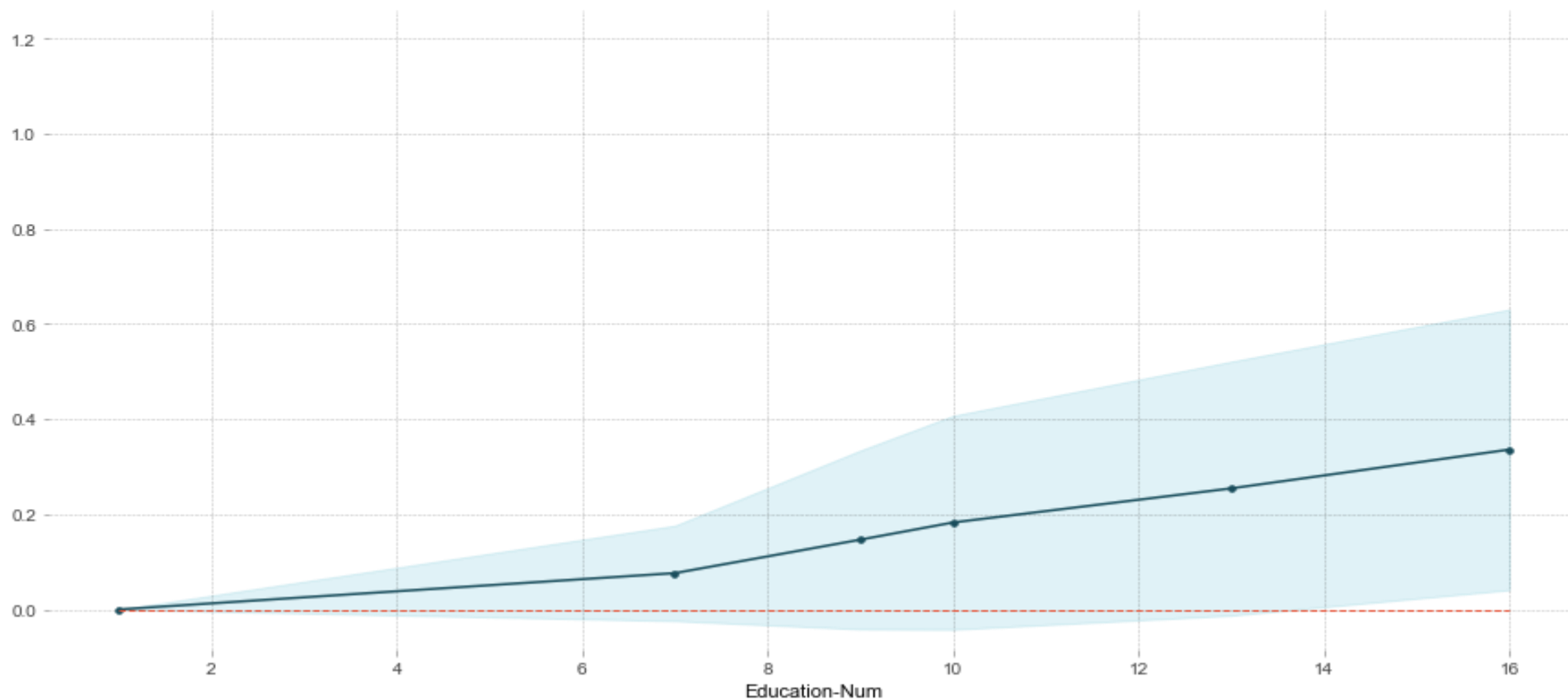
# ***Plots de Dependência Parcial***

## ***Univariado***

PDP for feature "Education-Num"

Number of unique grid points: 6

*lib Python:*  
***pdpbox***



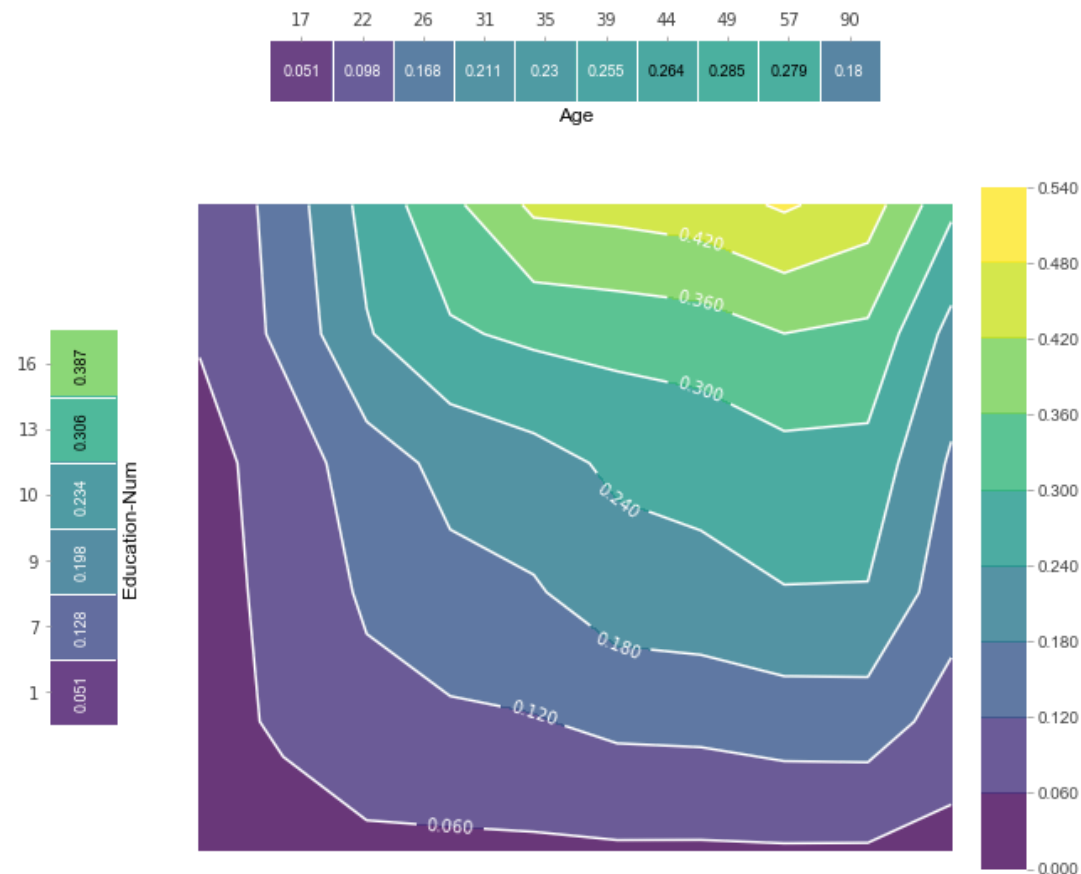
# ***Plots de Dependência Parcial***

## ***Bivariado***

*lib Python:*  
***pdpbox***

PDP interact for "Age" and "Education-Num"

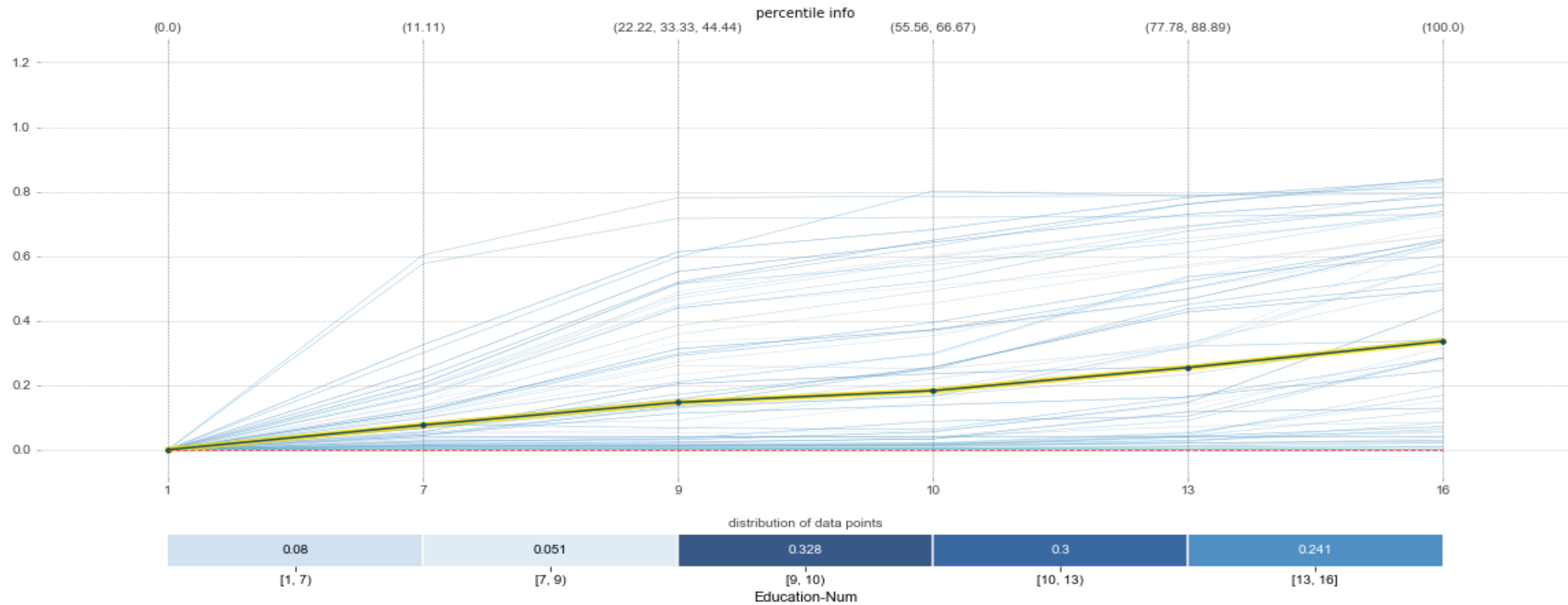
Number of unique grid points: (Age: 10, Education-Num: 6)





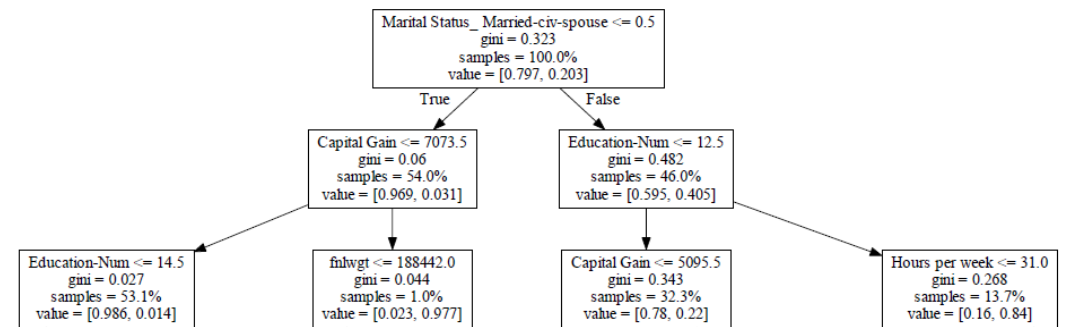
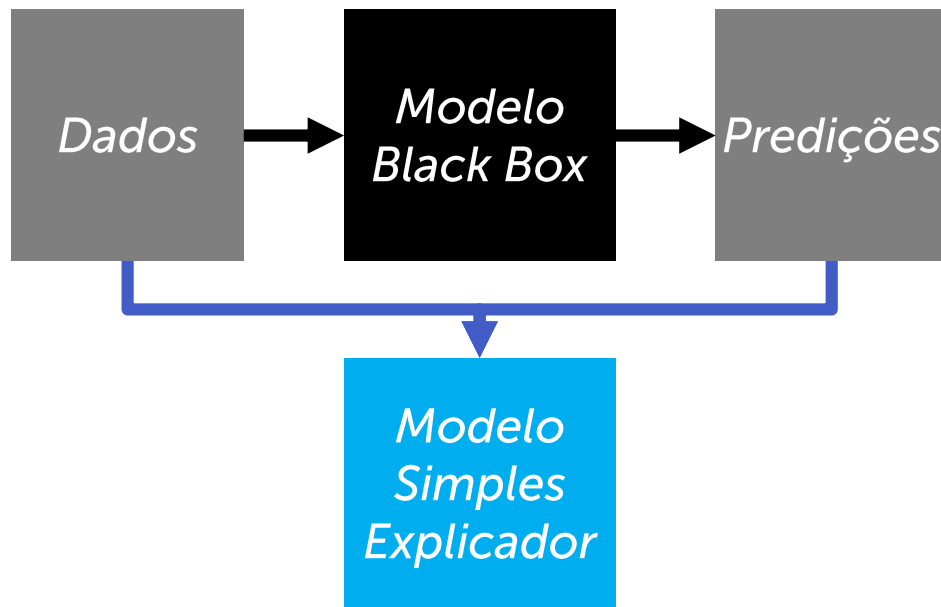
# Individual Conditional Expectation (ICE)

lib Python:  
**pdpbox**



# Global Surrogate Model

- Com as variáveis de entrada, cria-se um **modelo simples para prever as respostas do modelo complexo**.



Árvore de decisão simples gerada a partir de scores gerados por um Gradient Boosting

# ***Técnicas de interpretabilidade***

## **1. Técnicas de análises exploratórias**

- *Dependências parciais*
- *Individual Conditional Expectation (ICE)*
- *Global Surrogate Models*

## **2. Pacotes de interpretabilidade de modelos (Explainers)**

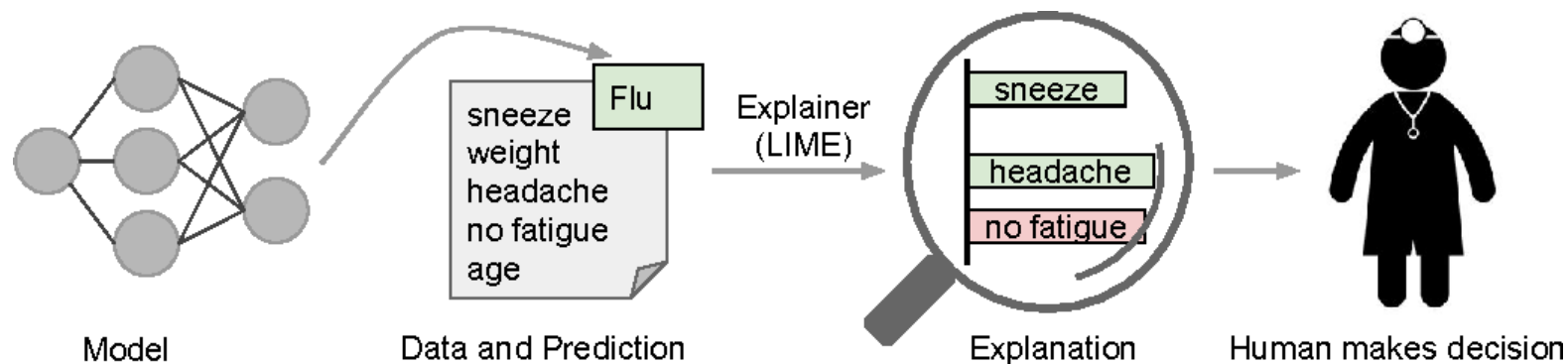
- *LIME*
- *SHAP*
- *eli5*



**CODE DEMO**

# LIME

- **Local Interpretable Model-Agnostic Explanations**
- É capaz de fazer **interpretações locais**, ao treinar um modelo linear na vizinhança do ponto de interesse.
- Implementações em Python (lime, Skater) e R (lime package, iml package), de fácil utilização



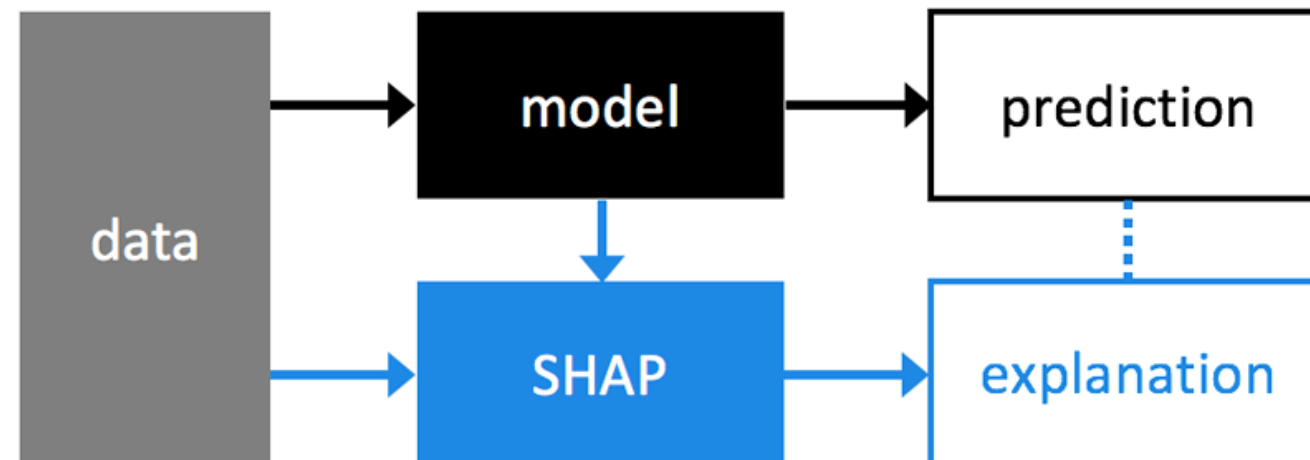
Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Why should I trust you?: Explaining the predictions of any classifier." (2016)


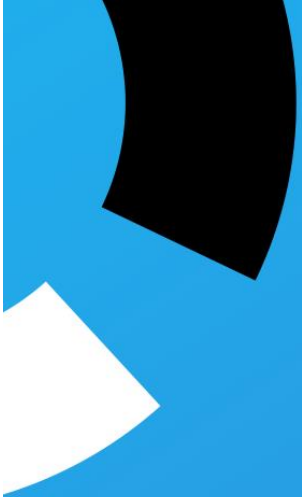
A large, faint, stylized graphic of a lime wheel, showing segments and a central core, is centered in the background.

# **DEMO LIME**

# **SHAP**

- **SH**apley **A**dditive ex**P**lanations
- Método tem origem em **Teoria dos Jogos**
- Implementações em Python (shap) e R (iml package)
- Calcula a **contribuição marginal média de uma variável** em todas as possíveis combinações de outras variáveis





# DEMO SHAP



# \_ Comparação SHAP vs LIME

## Pros

## Cons

### LIME

- *Explicação local simples e visual*
- *Execução rápida*

- *Conceito de “vizinhança” difícil de ser definido em problemas com variáveis categóricas*
- *Aproximação linear local pode não ser exata (verifique o  $R^2$ !)*
- *Não faz explicação global*

### SHAP

- *Explicação local e global*
- *Propriedades matemáticas fortes – Explicações sólidas, aderentes a órgãos reguladores.*

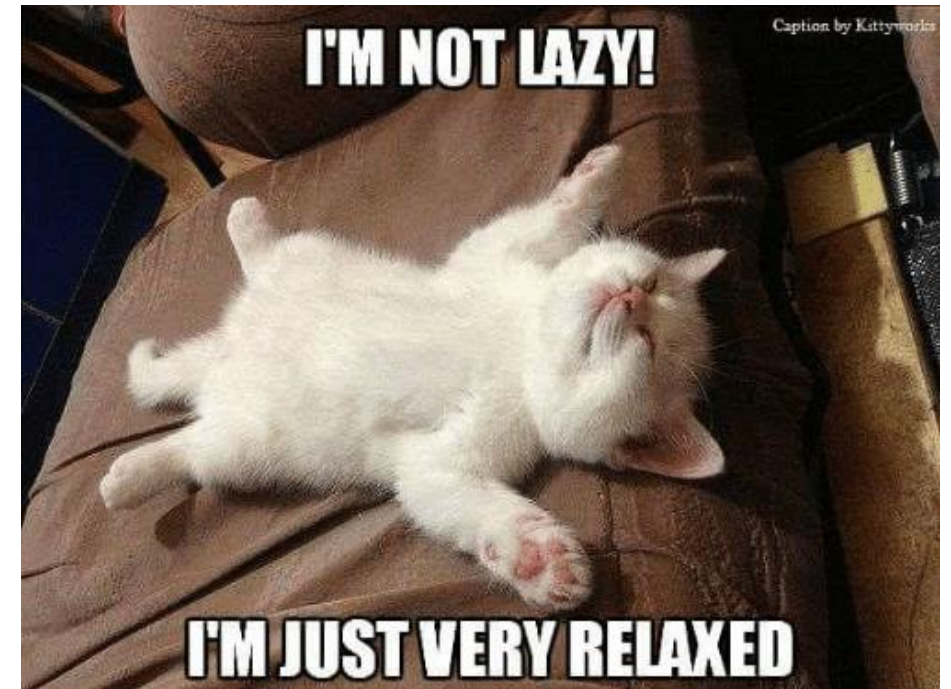
- *Alto tempo de processamento, devido ao cálculo de todas as combinações de features (otimizadores atenuam esse problema)*

## ***\_eli5: Para evitar a fadiga***

- ***Explain Like I'm 5***
- *API única em Python, muito simples de usar*
- *Funciona com scikit-learn, XGBoost, lightGBM...*

```
import eli5  
eli5.show_weights(algo)
```

Weight	Feature
0.3564	Marital_Status__Married-civ-spouse
0.1847	Capital_Gain
0.1553	Education-Num
0.0776	Age
0.0596	Capital_Loss
0.0468	Hours_per_week



# **Referências**

- **Christoph Molnar - Interpretable Machine Learning: A Guide for Making Black Box Models Explainable**  
<https://christophm.github.io/interpretable-ml-book/>
- **Scott Lundberg, Su-In Lee, "A unified approach to interpreting model predictions"**  
<https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>
- **Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Why should I trust you?: Explaining the predictions of any classifier."**  
<https://arxiv.org/abs/1602.04938>



# Data Scientist



# Coordenador de DS



## ***We're Hiring!***

<https://cielo.gupy.io>

cielo

Obrigado!



Cielo Data  
Science



Nos Ajudem a  
Melhorar!

