PALESTRANTES

JAYME ANCHANTE

Tendências e Oportunidades em Inteligência Artificial

PyPOA

9 de novembro, PUC, Porto Alegre

PyPOA 2019

09/11/2019 início às 8:00
PUC - Prédio 32
Salas 516 e 517

APOIO

ESCOLA POLITÉCNICA

PUCRS | IDEIA - Centro de Apoio ao Desenvolvimento Científico e Tecnológico

Tendências e Oportunidades em Inteligência Artificial

Jayme Anchante Linkedin | Github

Bacharel e mestre em Economia. Cientista de dados na Iris/4all. Pythonista. Ciclista, yogi e climber nas horas vagas

Sobre palestra: História do Deep Learning. Principais cases recentes que mudaram a história da área. Breves técnicas e teoria. Live coding com Python, Pytorch e Fastai

#RECRUTA

IRIS

Analista de Dados JR
Dev Backend Python
Dev Engenheiro de Dados
Dev Backend Node
Scrum Master
Dev Mobile Android
Dev Frontend
Dev Tester
Estágio CS

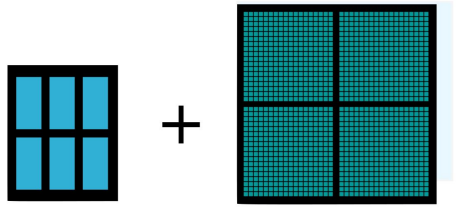
bit.ly/RecrutaIris

https://www.forbes.com/sites/janakirammsv/2018/05/27/here-are-three-factors-that-accelerate-the-rise-of-artificial-intelligence/#490aeb14add9

13,945 views | May 27, 2018, 10:59pm

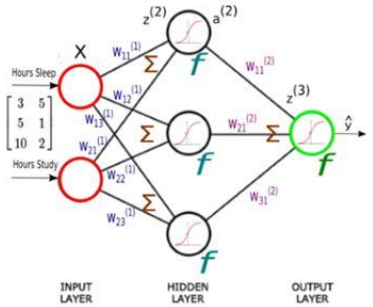
Here Are Three Factors That Accelerate The Rise Of Artificial Intelligence

Janakiram MSV Contributor ⓘ
Enterprise & Cloud
I cover Cloud Computing, Machine Learning, and Internet of Things



CPU
Multiple Cores

GPU
Thousands of Cores

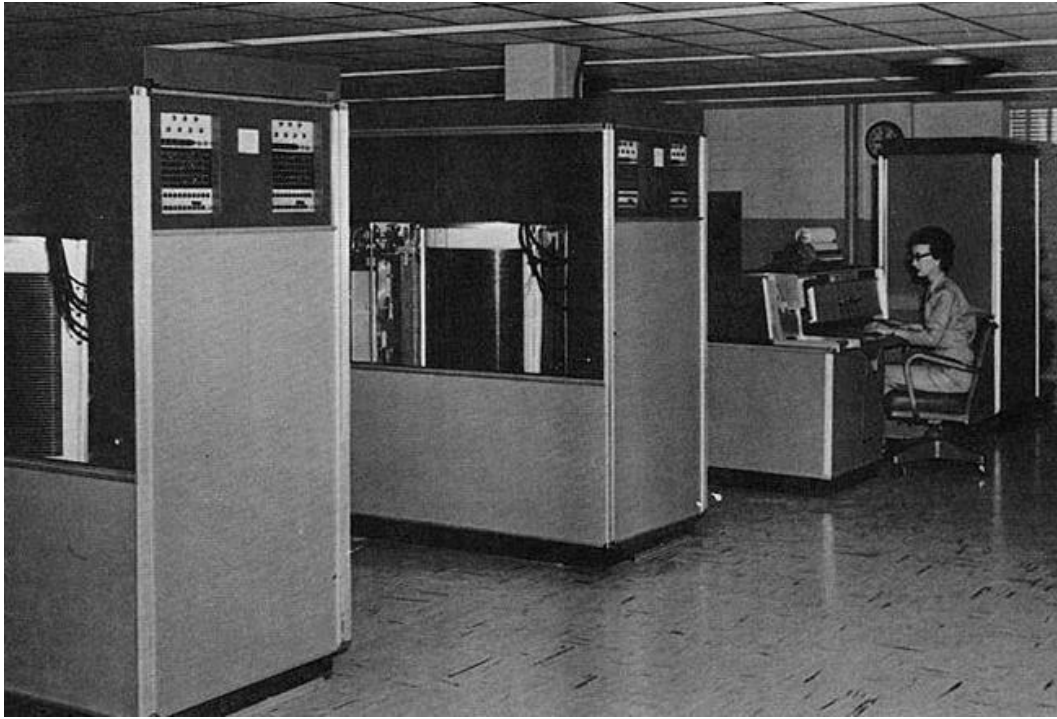


Prazer, Big Data

9m x 15m, mais de 1 tonelada

$ 3.200/mês o aluguel

5 MB de capacidade



RAMAC 305

1956: IBM lança o primeiro computador com disco rígido

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

https://www.dailymail.co.uk/sciencetech/article-2381188/Revealed-happens-just-ONE-minute-internet-216-000-photos-posted-278-000-Tweets-1-8m-Facebook-likes.html

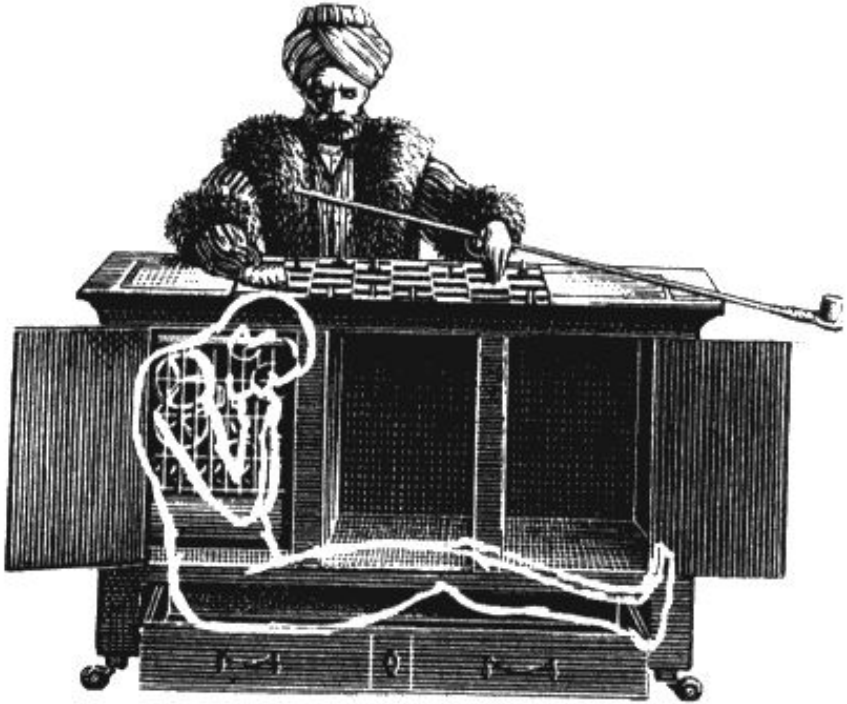


Plate 3.

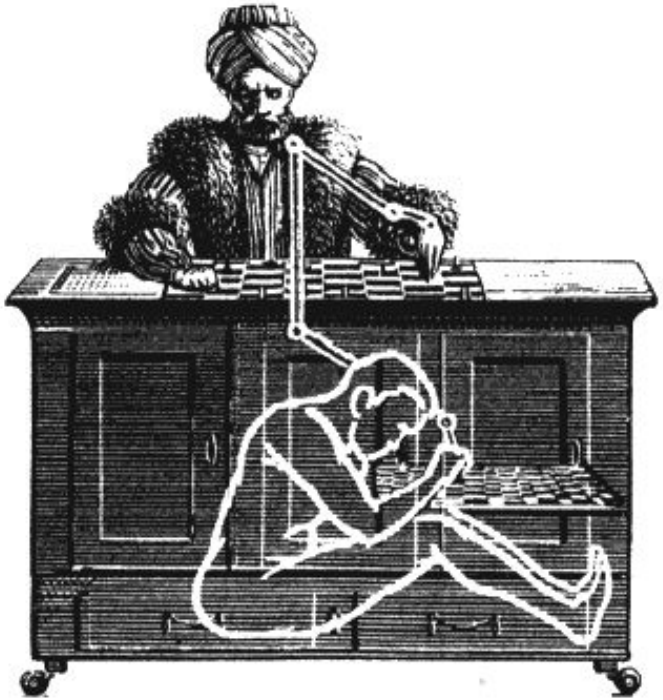


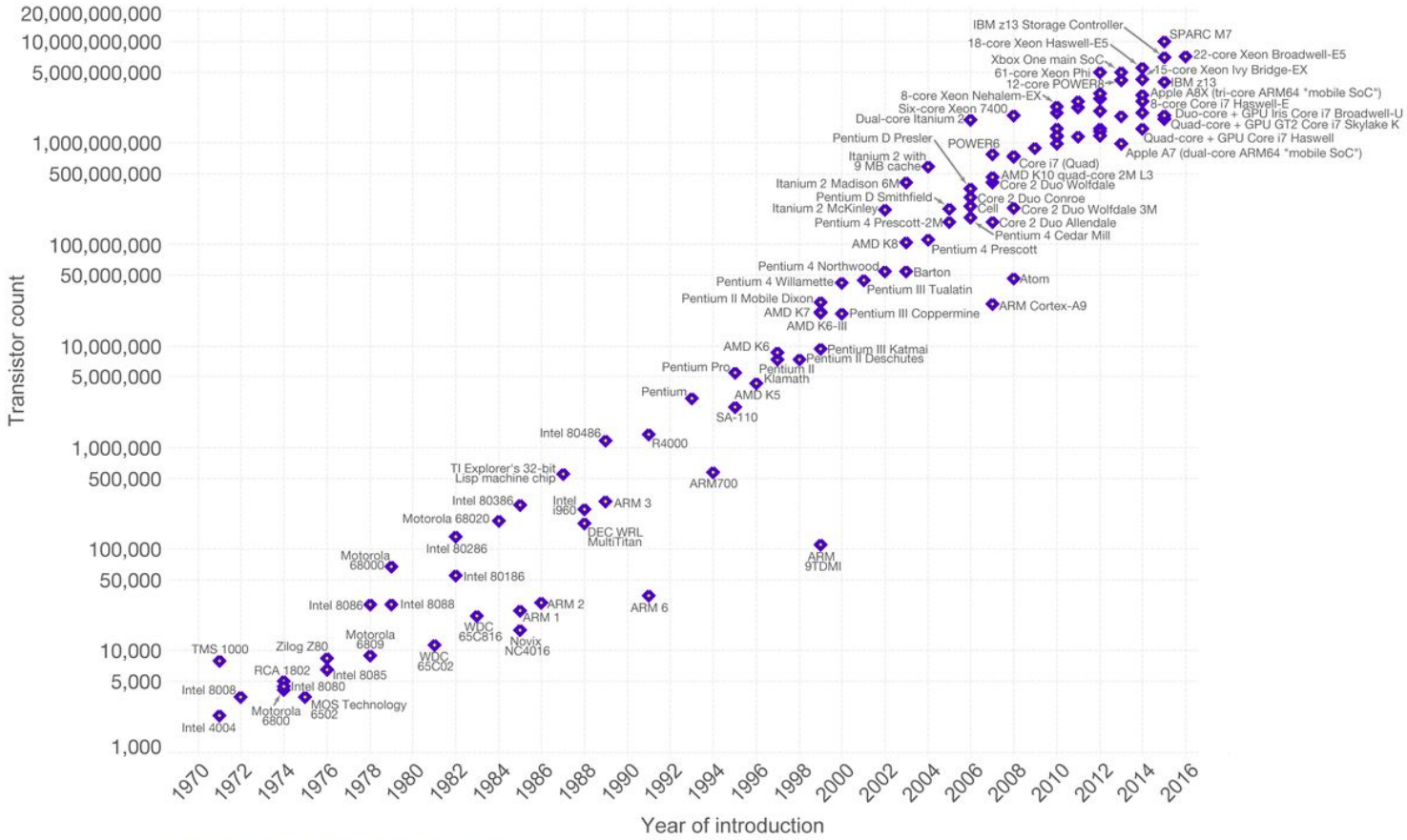
Plate 4.

https://eagereyes.org/wp-content/uploads/2016/05/mturk.png

Moore's Law – The number of transistors on integrated circuit chips (1971-2016)

Our World in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are strongly linked to Moore's law.



Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)

The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

Número de transistors em circuitos dobra a cada 2 anos

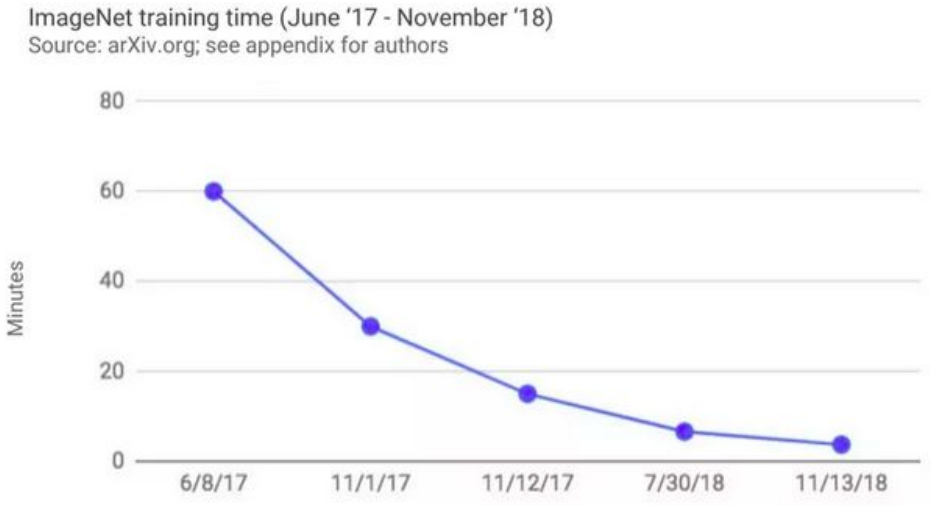
ImageNet Dataset

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *arXiv preprint arXiv:1409.0575*. [web]

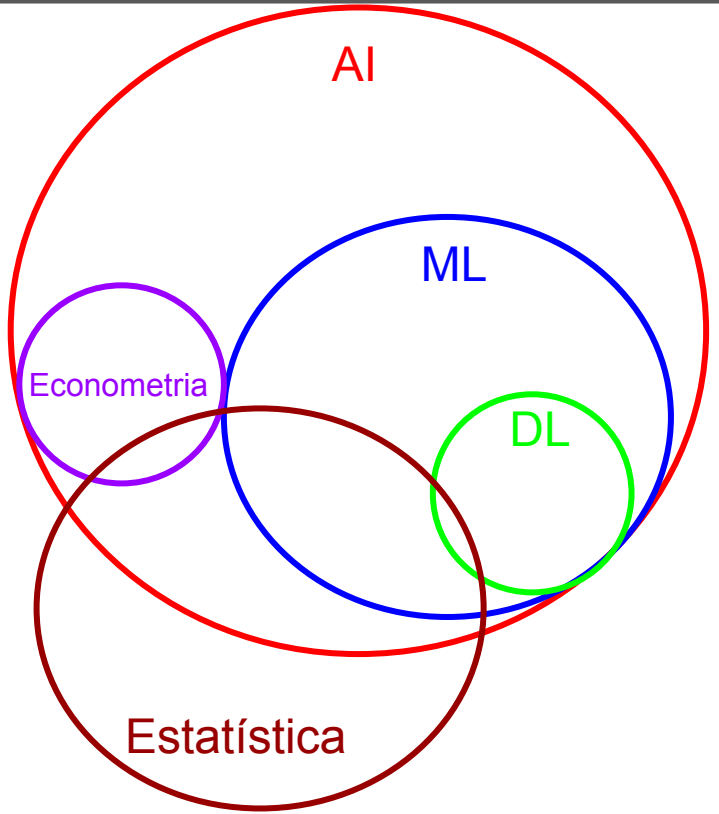
3.2M imagens

5247 classes

2.5 anos para rotular



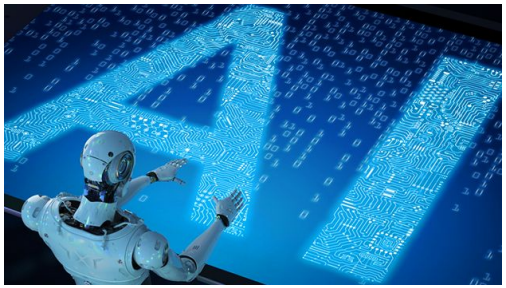
mundo fora das buzzwords



Inteligência Artificial

Inteligência demonstrada pelas máquinas

Tarefas usuais: entender a fala humana, competir em jogos de estratégia, carros autônomos, rede de entrega de conteúdos, simulações militares



Inteligência Natural

Inteligência demonstrado por humanos e outros animais

Aprendizado de máquina

Campo da ciências da computação que utiliza técnicas estatísticas (**erro**) para dar a computadores a habilidade de aprender com os dados, sem qualquer programação explícita

Recomendação de produto

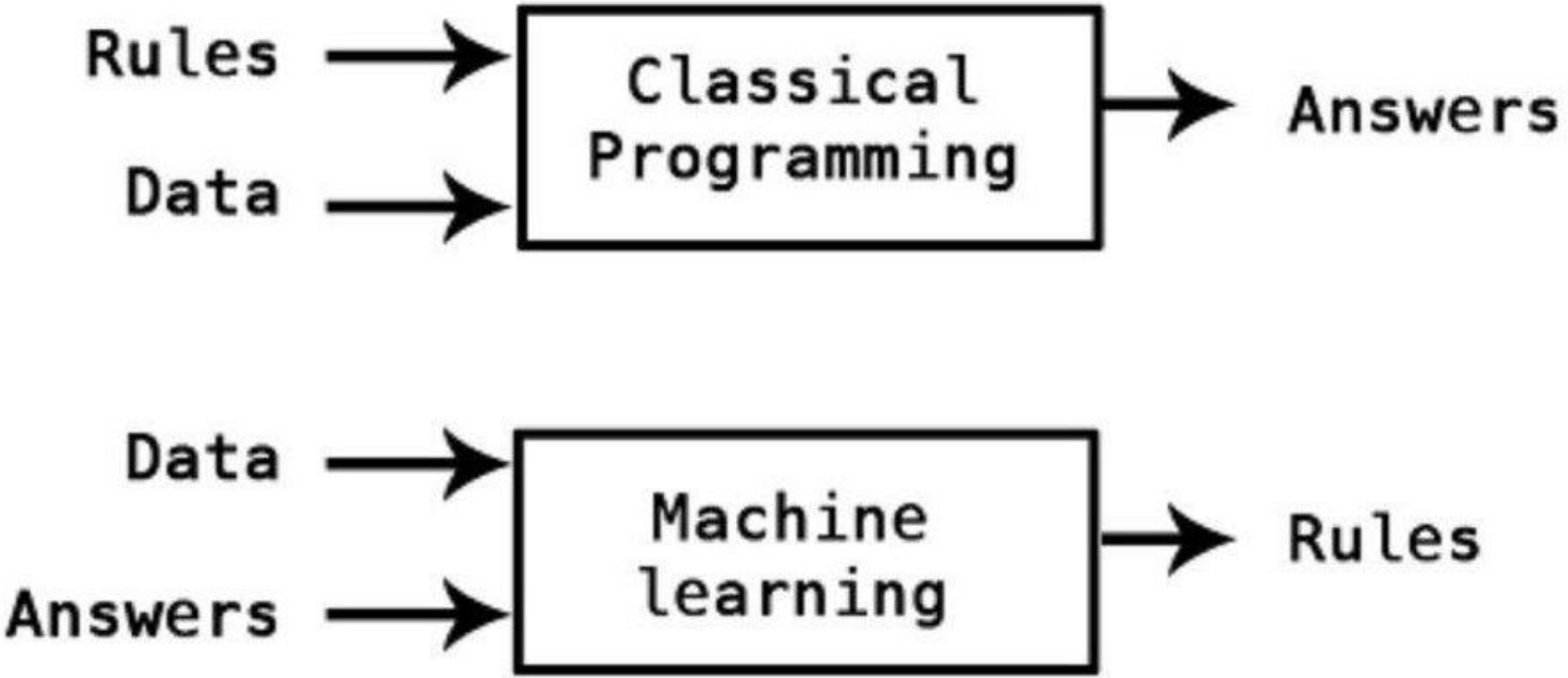
1 Ajustar uma regressão logística com base no histórico de compra, idade e gênero

Regras determinísticas

GOFAI ("Good Old-Fashioned Artificial Intelligence") ou Symbolic artificial intelligence são uma coleção de métodos de Inteligência Artificial baseados em representações simbólicas de alto nível, tais como lógica e busca. **Não envolve erro!**
(Foi dominante entre os anos 50 e fim dos 80)

Recomendação de produto

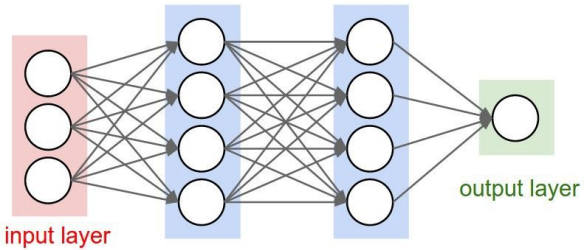
1 Se Idade > 30 anos e Gênero == Homem:
    Recomendar produto 1
2 Se Idade <= 30 anos e Gênero == Homem:
    Recomendar produto 2
3 Se Idade > 30 anos e Gênero == Mulher:
    Recomendar produto 3
4 Se Idade <= 30 anos e Gênero == Mulher:
    Recomendar produto 4



https://www.ben-evans.com/benedictevans/2019/9/6/face-recognition; François Chollet

Aprendizado profundo

Aprendizado “profundo”, as camadas escondidas das redes neurais agem fazem uma engenharia de característica automática, revelando correlações não lineares entre as características

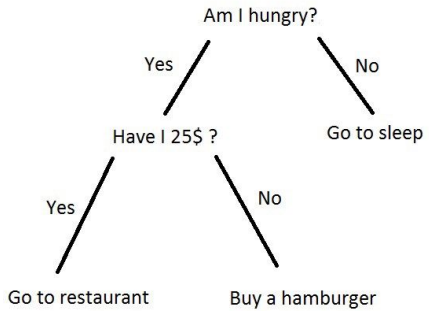


Correlações não-lineares

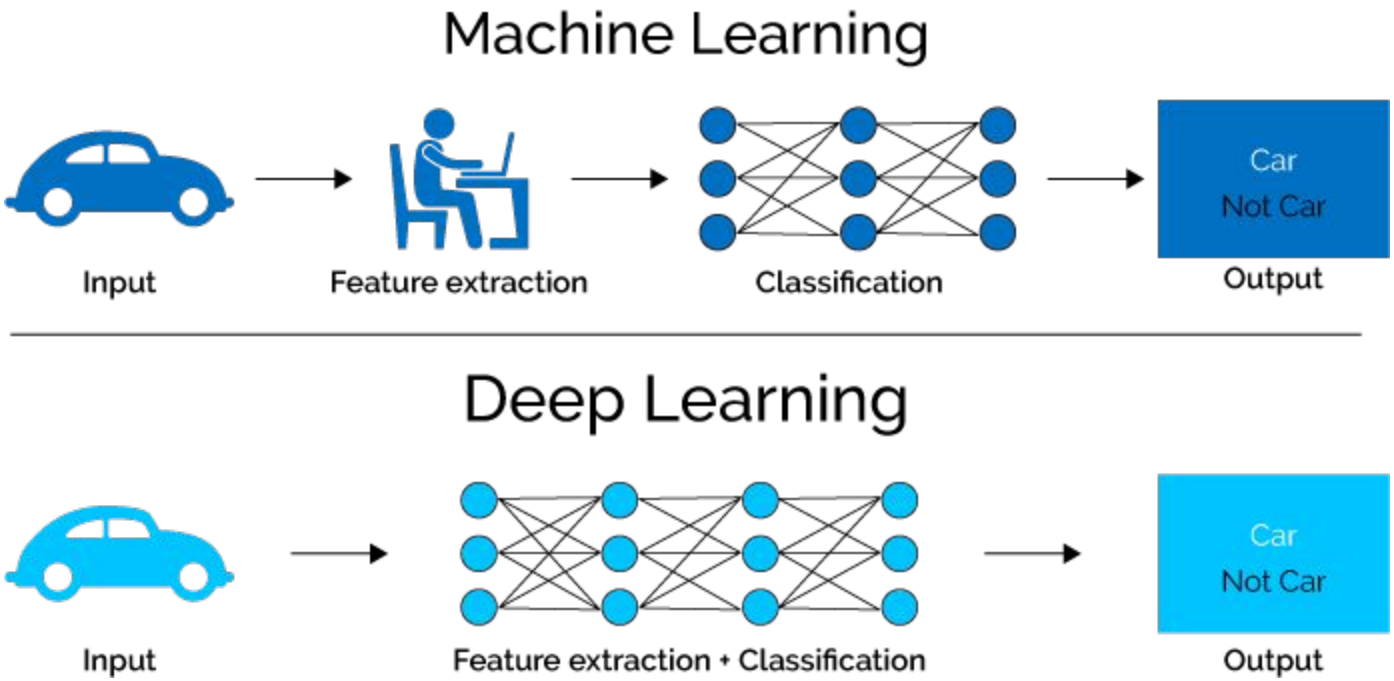
Exemplos: redes neurais profundas, recorrentes, convolucionais

Aprendizado raso

Aprendizado “raso”, a engenharia de características é feita fora e antes do algoritmo. A árvore de decisão abaixo é capaz de separar/classificar as pessoas, mas ela não é capaz de criar novas características



Exemplos: árvore de decisão, florestas aleatórias, regressão linear/logística


Machine Learning
Input
Feature extraction
Classification
Car
Not Car
Output
Deep Learning
Input
Feature extraction + Classification
Car
Not Car
Output

11/05/1997

6 jogos contra Gary Kasparov

15/03/2016

5 jogos contra Lee Sedol

13/04/2019

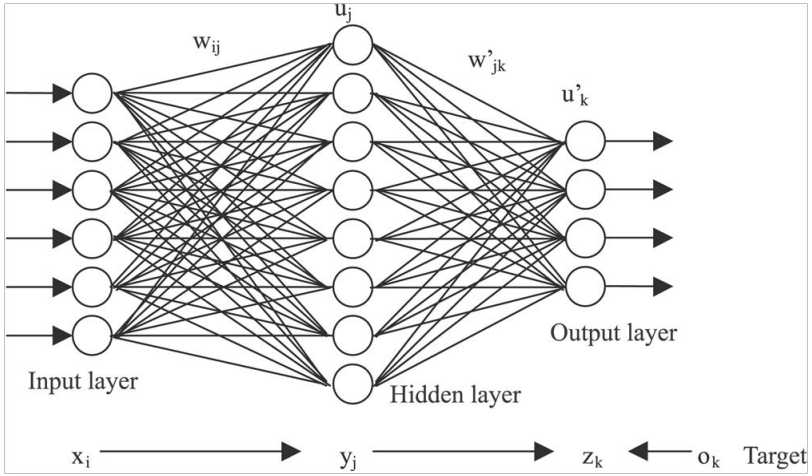
Campeonato mundial de Dota 2

Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An "Explainable AI" Competition

by Cynthia Rudin and Joanna Radin

Forthcoming. Read the abstract to learn more!

https://hdsr.mitpress.mit.edu/pub/f9kuryi8



https://www.extremetech.com/wp-content/uploads/2015/07/NeuralNetwork.png

Please Stop Explaining Black Box Models for High-Stakes Decisions

Cynthia Rudin
Duke University

ABSTRACT

Black box machine learning models are currently being used for high stakes decision-making throughout society, causing problems throughout healthcare, criminal justice, and in other domains. People have hoped that creating methods for explaining these black box models will alleviate some of these problems, but trying to *explain* black box models, rather than creating models that are *interpretable* in the first place, is likely to perpetuate bad practices and can potentially cause catastrophic harm to society. There is a way forward – it is to design models that are inherently interpretable.

https://www.arxiv-vanity.com/papers/1811.10154/

```
import eli5
eli5.show_weights(clf, top=10)
```

y=0 top features		y=1 top features		y=2 top features		y=3 top features	
Weight[?]	Feature	Weight[?]	Feature	Weight[?]	Feature	Weight[?]	Feature
+1.991	x21167	+1.702	x15699	+2.016	x25234	+1.193	x28473
+1.925	x19218	+0.825	x17366	+1.951	x12026	+1.030	x8609
+1.834	x5714	+0.798	x14281	+1.758	x17854	+1.021	x8559
+1.813	x23677	+0.786	x30117	+1.697	x11729	+0.946	x8798
+1.697	x15511	+0.779	x14277	+1.655	x32847	+0.899	x8544
+1.696	x26415	+0.773	x17356	+1.522	x22379	+0.797	x8553
+1.617	x6440	+0.729	x24267	+1.518	x16328	*... 11122 more positive ...*	
+1.594	x26412	+0.724	x7874	*... 15007 more positive ...*		*... 24657 more negative ...*	
... 10174 more positive ...		+0.702	x2148	*... 20772 more negative ...*		-0.852	x15699
... 25605 more negative ...		*... 11710 more positive ...*		-1.764	x15521	-0.894	x25663
-1.686	x28473	*... 24069 more negative ...*		-2.171	x15699	-1.181	x23122
-10.453	<BIAS>	-1.379	<BIAS>	-5.013	<BIAS>	-1.243	x16881

https://eli5.readthedocs.io/en/latest/tutorials/sklearn-text.html

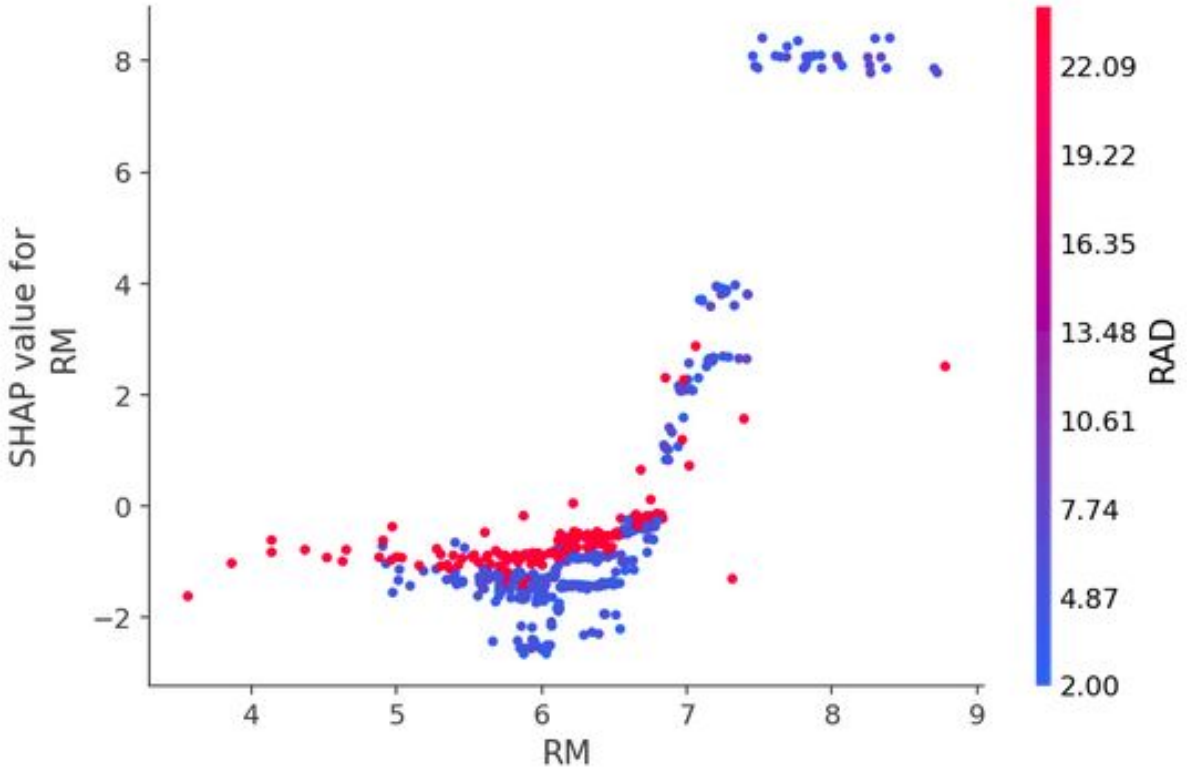
```
from eli5 import show_prediction
show_prediction(clf, valid_xs[1], vec=vec, show_feature_values=True)
```

y=1 (probability **0.566**, score **0.264**) top features

Contribution[?]	Feature	Value
+1.673	Sex=female	1.000
+0.479	Embarked=S	Missing
+0.070	Fare	7.879
-0.004	Cabin=	1.000
-0.006	Parch	0.000
-0.009	Pclass=2	Missing
-0.009	Ticket=1601	Missing
-0.012	Embarked=C	Missing
-0.071	SibSp	0.000
-0.073	Pclass=1	Missing
-0.147	Age	19.000
-0.528	<BIAS>	1.000
-1.100	Pclass=3	1.000

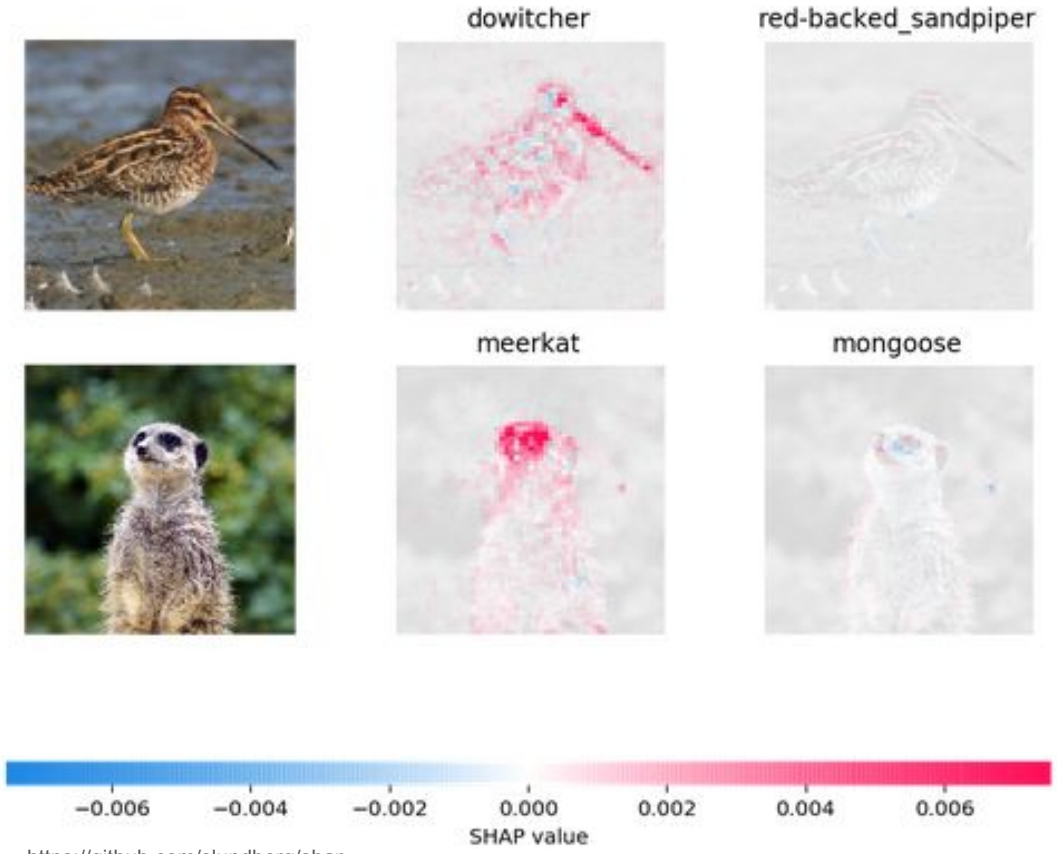
https://eli5.readthedocs.io/en/latest/tutorials/xgboost-titanic.html

```
# create a SHAP dependence plot to show the effect of a single feature across the whole dataset
shap.dependence_plot("RM", shap_values, X)
```



https://github.com/slundberg/shap

```
# plot the explanations
shap.image_plot(shap_values, to_explain, index_names)
```



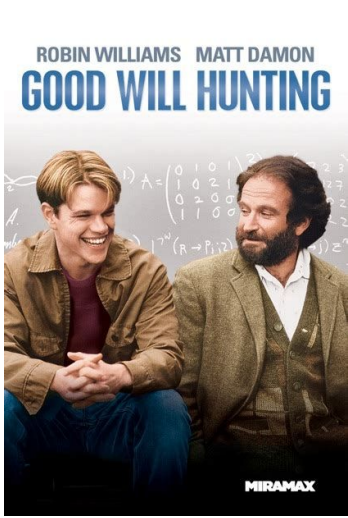
https://github.com/slundberg/shap

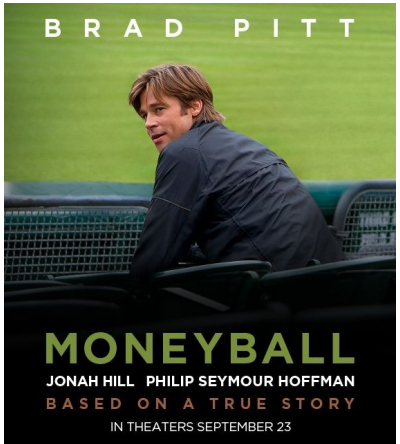

HIDDEN FIGURES
COMING SOON


INSPIRED BY A TRUE STORY
MARGIN CALL
BE FIRST.
BE SMARTER.
OR CHEAT.
COMING SOON

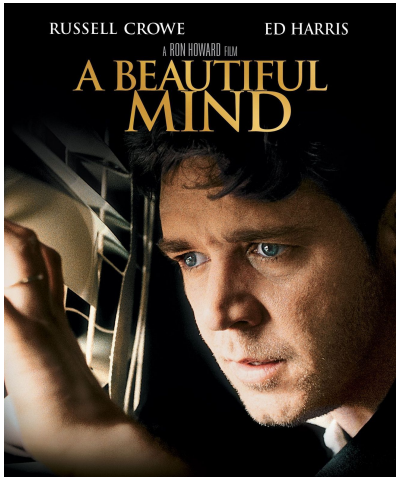

CHRISTIAN BALE
STEVE CARELL
RYAN GOSLING
BRAD PITT
THE BIG SHORT
THIS IS A TRUE STORY
FROM THE AUTHOR OF
'MONEYBALL'
AND
'THE BLIND SIDE'


THEY TOOK YOUR DATA.
THEN THEY TOOK CONTROL.
A NETFLIX ORIGINAL DOCUMENTARY
THE GREAT HACK
From the Academy Award® nominated filmmakers behind THE SQUARE
sundance
JULY 24 | NETFLIX


ROBIN WILLIAMS
MATT DAMON
GOOD WILL HUNTING
MIRAMAX


BRAD PITT
MONEYBALL
JONAH HILL PHILIP SEYMOUR HOFFMAN
BASED ON A TRUE STORY
IN THEATERS SEPTEMBER 23


NETFLIX
ALPHAGO
AlphaGo
Lee Sedol


RUSSELL CROWE
ED HARRIS
A RON HOWARD FILM
A BEAUTIFUL MIND

Case de ciência de dados

Aplicabilidade e usos para negócios

Visualização

Importância de divisão da base de treino-validação-teste

Explicabilidade de um modelo

Interpretação

Performance

UCI Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Restaurant & consumer data Data Set

Download: Data Folder, Data Set Description

Abstract: The dataset was obtained from a recommender system prototype. The task was to generate a top-n list of restaurants according to the consumer preferences.

Data Set Characteristics:	Multivariate	**Number of Instances:**	138	**Area:**	Computer
Attribute Characteristics:	N/A	**Number of Attributes:**	47	**Date Donated**	2012-08-04
Associated Tasks:	N/A	**Missing Values?**	Yes	**Number of Web Hits:**	128282

Source:

Creators:
Rafael Ponce MedellÃn and Juan Gabriel GonzÃ¡lez Serna
rafaponce '@' cenidet.edu.mx, gabriel '@' cenidet.edu.mx
Department of Computer Science.
National Center for Research and Technological Development CENIDET, MÃ©xico
Donors of database:
Blanca Vargas-Govea and Juan Gabriel GonzÃ¡lez Serna
blanca.vargas '@' cenidet.edu.mx/blanca.vg@gmail.com, gabriel '@' cenidet.edu.mx
Department of Computer Science.
National Center for Research and Technological Development CENIDET, MÃ©xico

Data Set Information:

Two approaches were tested: a collaborative filter technique and a contextual approach.
(i) The collaborative filter technique used only one file i.e., rating_final.csv that comprises the user, item and rating attributes.
(ii) The contextual approach generated the recommendations using the remaining eight data files.

Attribute Information:

Files, instances and attributes
Number of Files: 9

Restaurants
1 chefmozaccepts.csv
2 chefmozcuisine.csv
3 chefmozhours4.csv
4 chefmozparking.csv
5 geoplaces2.csv

https://archive.ics.uci.edu/ml/datasets/Restaurant+%26+consumer+data