

Online

# Oficina Gratuita Machine Learning



AI New Ventures

31/10 - 14h às 17h  
Transmissão via [Youtube](#).

Instituição beneficiada

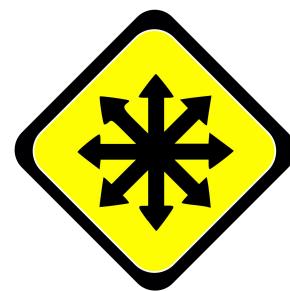


*Instituto  
do Câncer  
Infantil*

Começaremos em instantes...

# Edições Anteriores

(+1.000 participantes)





Frederico Tarrago • 1º  
Inovação e Processos e Coordenador da P...  
1 m

Oficina/Aula de altíssimo padrão.  
Recomendo muito! Parabéns pela  
iniciativa [Henrique Dias!](#)!



Douglas Trajano • 1º  
Analista integrador pleno | Sicredi  
1 m

O [Henrique Dias](#) é fantástico. Ele explica  
de forma simples e divertida. :)



Israel G. de Oliveira • 1º  
Data Scientist : ADP Brazil Labs | Instructor...  
2 d

Recomendo fortemente! Excelente oficina, eu fui e foi  
muito produtivo e divertido. O Henrique e os demais  
instrutores fazem o tempo passar rápido.



Nana Raythz (Natalia) • 2º  
Data Engineer | DevOps | PCD | DEPC | SFP...  
1 d  
O [Henrique Dias](#) é demais



Caesar de Cesar • 1º  
Desenvolvedor Web JAVA Sênior na...  
1 h • Editado

Eu não sou a Betina, mas tenho uma dica de mega  
investimento: São só 15 reais para um café, mas  
que podem mudar tua vida, já fui em uma edição  
passada e nunca vou deixar de recomendar, não só  
pelo conteúdo, mas principalmente pela dinâmica e  
as inspirações!



Vinícius Marchi Appel • 1º  
Especialização Ciência de Dados | PUCRS e...  
2 m

Valeu muito esse treinamento! Obrigado, [Henrique Dias!](#)



Marcio Almeida • 1º  
Data Engineer ETL BI Cloud Computing Big D...  
1 d

Ótimo evento !!  
Recomendo !!



Douglas Trajano • 1º  
Analista integrador pleno | Sicredi  
1 d

Super recomendo a oficina de I.A. ministrado pelo  
[Henrique Dias!](#) O cara é bom demais e explica de uma  
forma muito divertida e simples sobre como usar ML  
na prática.

# Público-Alvo

---



Devs



Linguistas



Jornalistas



Economistas



Engenheirxs



Advogadxs



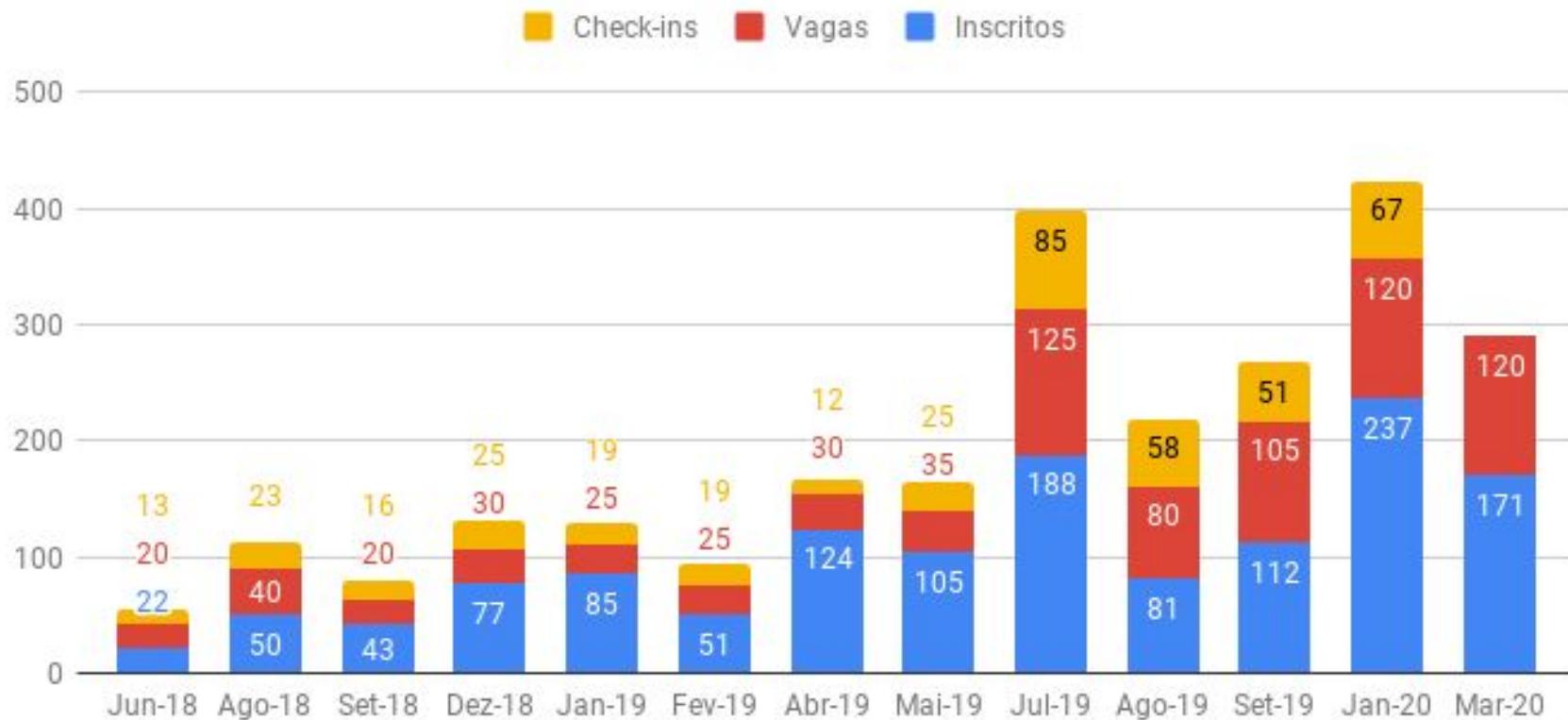
Prof. Saúde



Estudante

...

## Inscritos, Vagas and Check-ins



## DETALHES DO EVENTO

Status Visualizações da página do evento  
● **Publicado** 3.489

Visibilidade Compartilhar o evento    COPIAR ENDEREÇO  
Público ●  
[https://www.sympla.com.br/oficina-gratuita-de-machine-learning-24-edicao--navi\\_77...](https://www.sympla.com.br/oficina-gratuita-de-machine-learning-24-edicao--navi_77...)

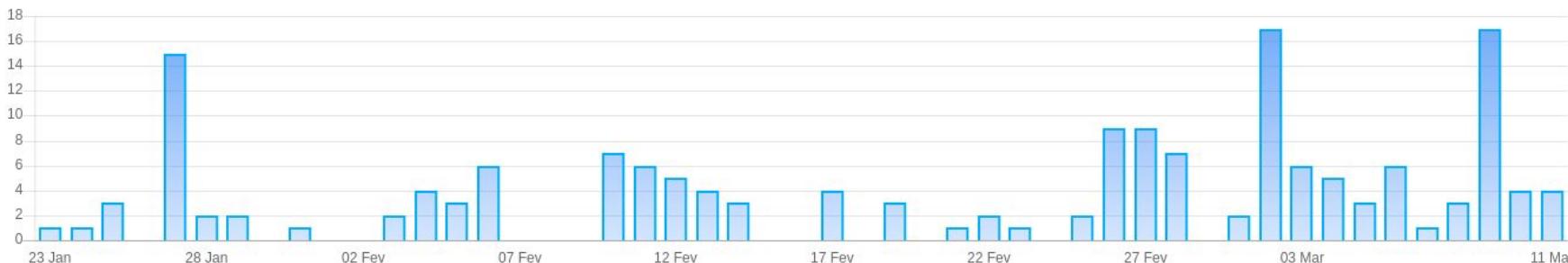
## FINANCEIRO

Vendas totais **R\$ 2.400,00** Em processamento (pendentes) **R\$ 0,00**  
Total a receber ● **R\$ 2.100,00** Coffee-break **R\$ 2.053,00**

INGRESSOS ●

Período: Filtrar por canal de venda: Filtrar por valor do ingresso:

Desde a publicação ▼ Todos os canais ▼ Gratuito e pago ▼ Todos os tipos ▼



Ingressos confirmados

**171**

Ingressos pendentes

**0**Ingressos cancelados ●**0**

# Check-In Sympla

---



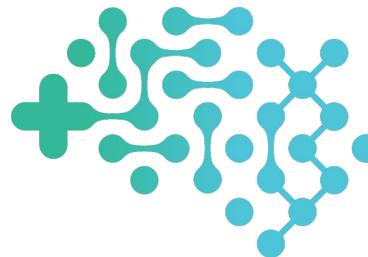
para emissão dos certificados

# Facilitador

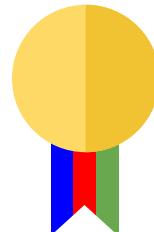
---



Henrique



PUCRS  
Grupo de  
**Inteligência Artificial**  
na Saúde



Google  
Research  
Award  
2018, 2019



# Introdução

# Data Science: The Movie

---

**DATA**<sup>PIONEERS</sup>  
**SCIENCE**

<https://vimeo.com/352483111>

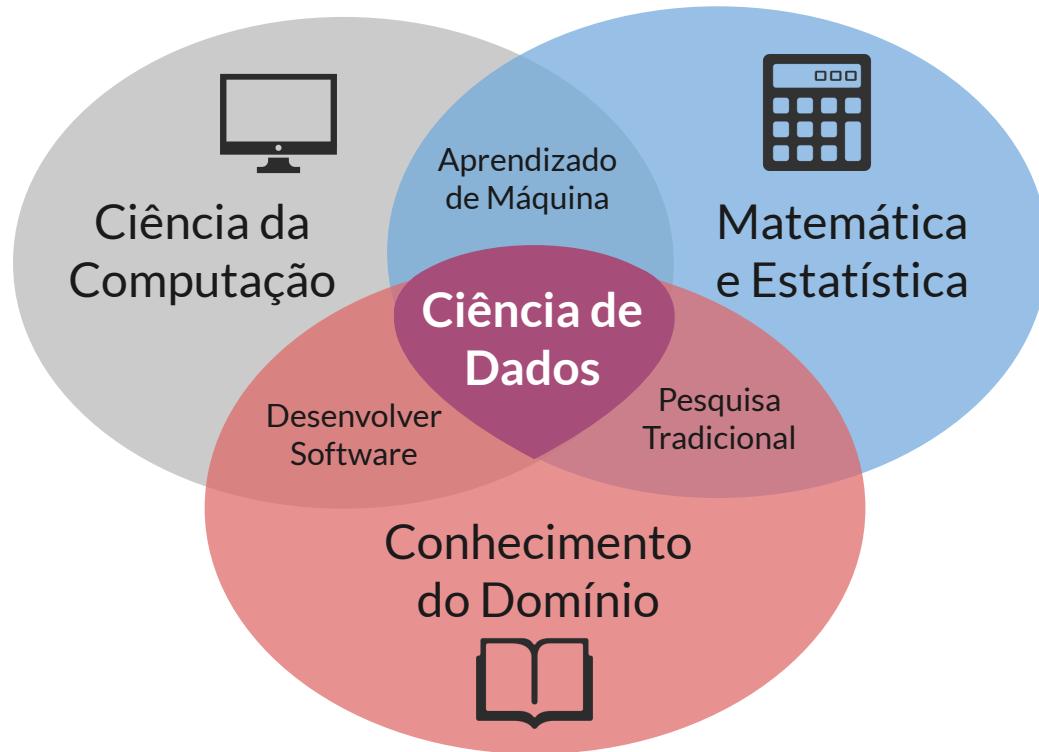
# Sopa de Letrinhas

---

Descoberta do Conhecimento      Computação Cognitiva  
Inteligência Artificial      Dados      Banco de Dados  
Computação em Núvem      Redes Neurais  
Aprendizado      Ciência de Dados  
Big Data      Padrões      de Máquina  
Supervisionada      Mineração de Dados      Inteligência Computacional  
Computação Instintiva      Indústria 4.0

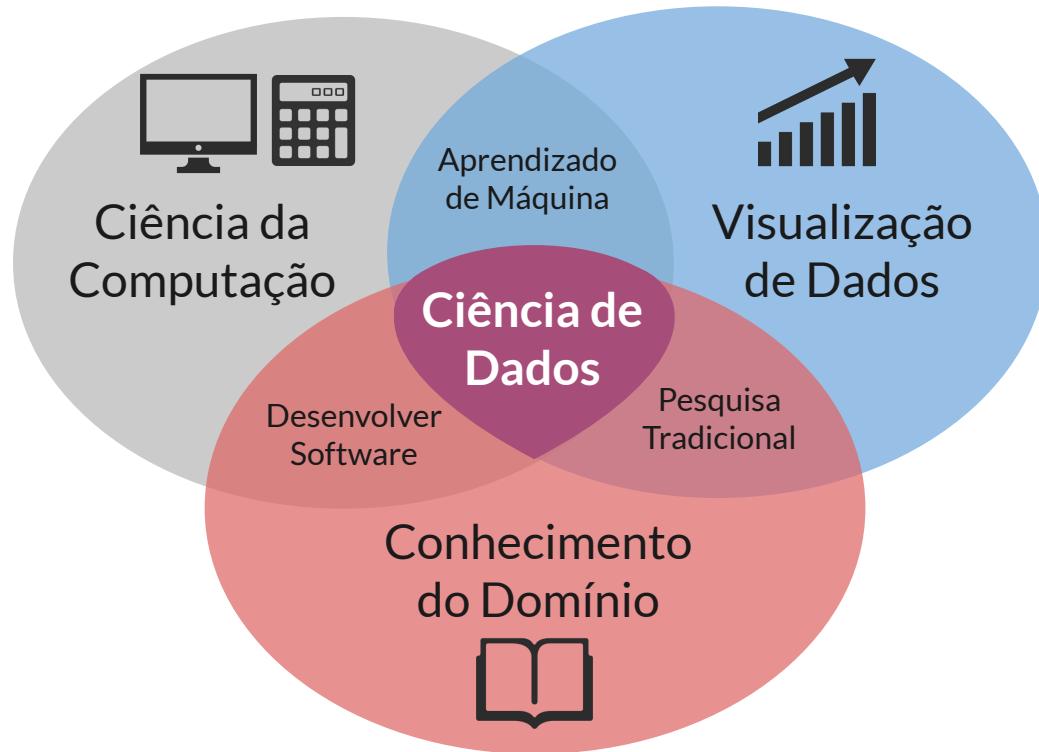
# Ciência de Dados

---



# Ciência de Dados

---



# MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

# DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business



# PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing package e.g. R
- ★ Databases SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with xaaS like AWS

# COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior

## ENGAGEMENT PROCESS

Step 1: Build the Data Model



Step 2: Define The Report



Step 3: Generate SQL commands



Step 4: Create Report



The data warehouse is a "schema-on-load" approach because the data schema must be defined and built prior to loading data into the data warehouse. Without an underlying data model, the BI tools will not work.

### Business Intelligence Questions

What happened?

Descriptive Analysis  
Standard Reporting



Business Intelligence Analyst

### Data Science Questions

Why? What will happen?  
What should I do?

Data Scientist

Predictive Analysis  
Prescriptive Analysis



## DIFFERENCE BETWEEN BUSINESS INTELLIGENCE AND DATA SCIENCE

### CHARACTERISTICS

<b>Focus</b>	Reports, KPIs, trends	Patterns, correlations, models
<b>Process</b>	Static, comparative	Exploratory, experimentation, visual
<b>Data Sources</b>	Pre-planned, added slowly	On the fly, as-needed
<b>Transform</b>	Up front, carefully planned	In-database, on-demand, enrichment
<b>Data quality</b>	Single version of truth	"Good enough," probabilities
<b>Data model</b>	Schema on load	Schema on query
<b>Analytics</b>	Retrospective, Descriptive	Predictive, Prescriptive, Preventative

## ENGAGEMENT PROCESS

Step 1: Define Hypothesis to Test ←



Step 2: Gather Data



Step 3: Build Data Model



repeat

Step 4: Explore the Data



Step 5: Build and Refine Analytic Models



Step 6: Ascertain Goodness of Fit



# DATA Engineer

Develops, constructs, tests,  
and maintains architectures.  
Such as databases  
and large-scale  
processing systems.



DataCamp  
Learn Data Science By Doing

# DATA Scientist

Cleans, massages  
and organizes (big) data.  
Performs descriptive statistics  
and analysis to develop  
insights, build models and  
solve a business need.



# Ferramentas

---

## Machine Learning



## Big Data



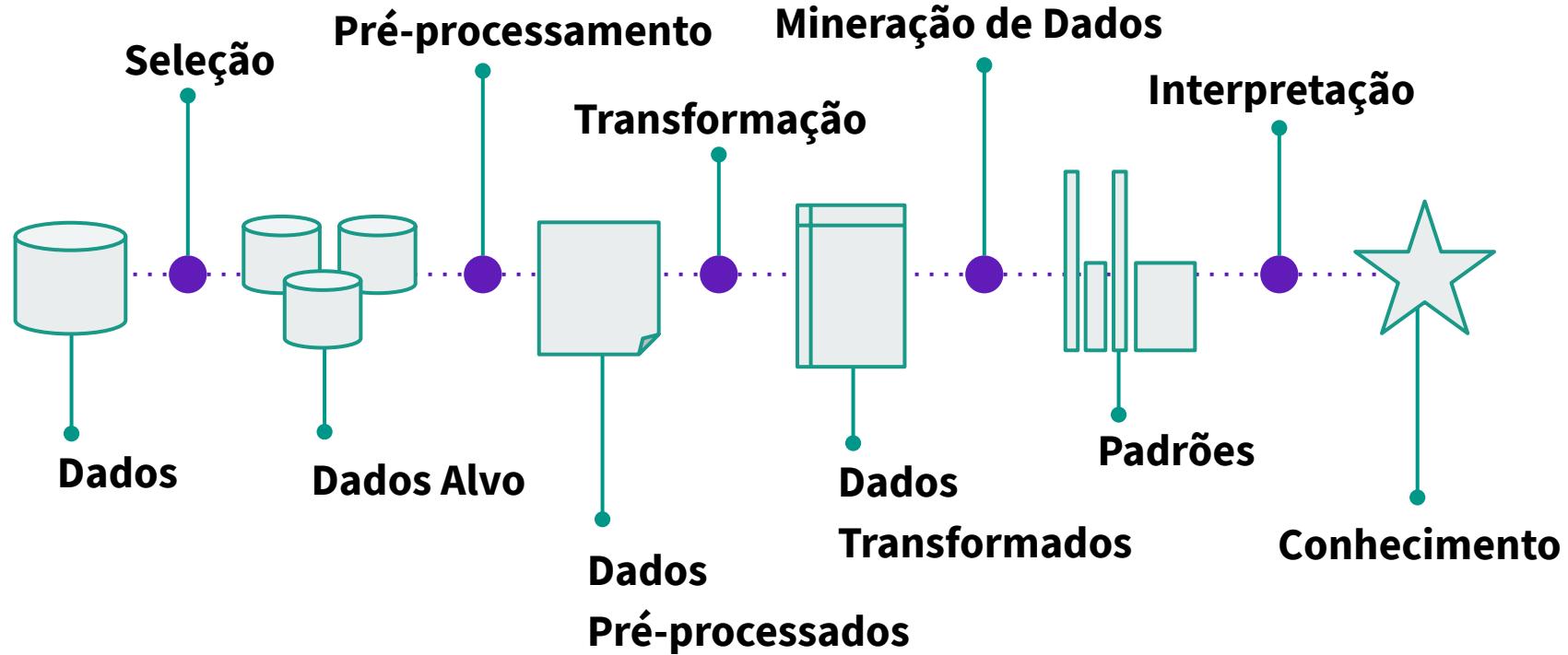
# Ferramentas

## Gartner Magic Quadrant Data Science

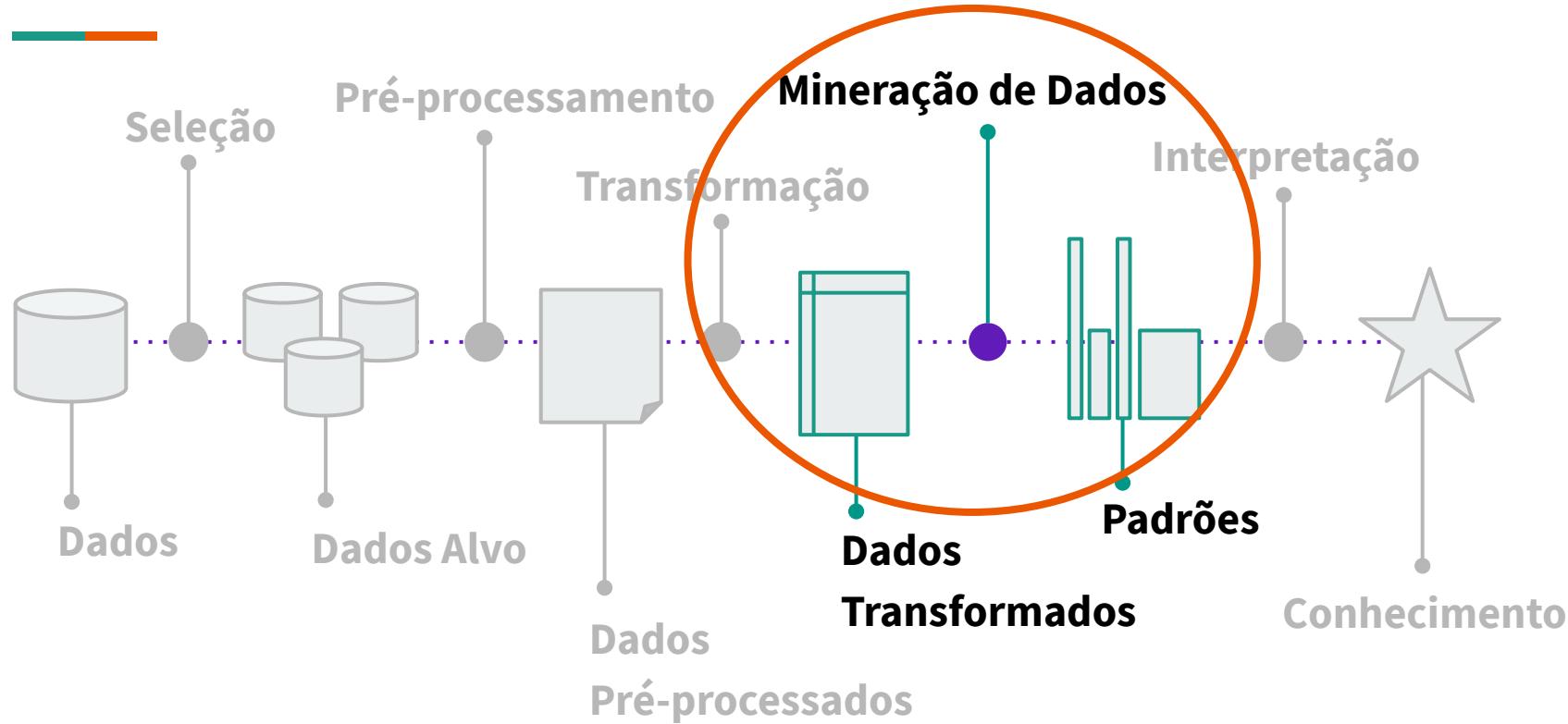
Figure 1. Magic Quadrant for Data Science and Machine Learning Platforms



# Descoberta de Conhecimento

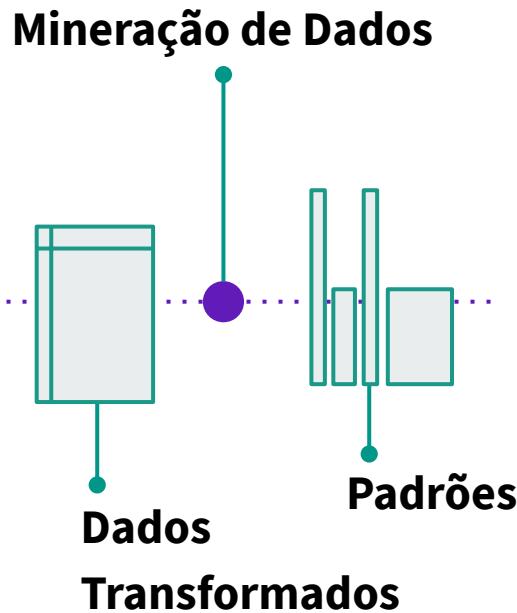


# Descoberta de Conhecimento



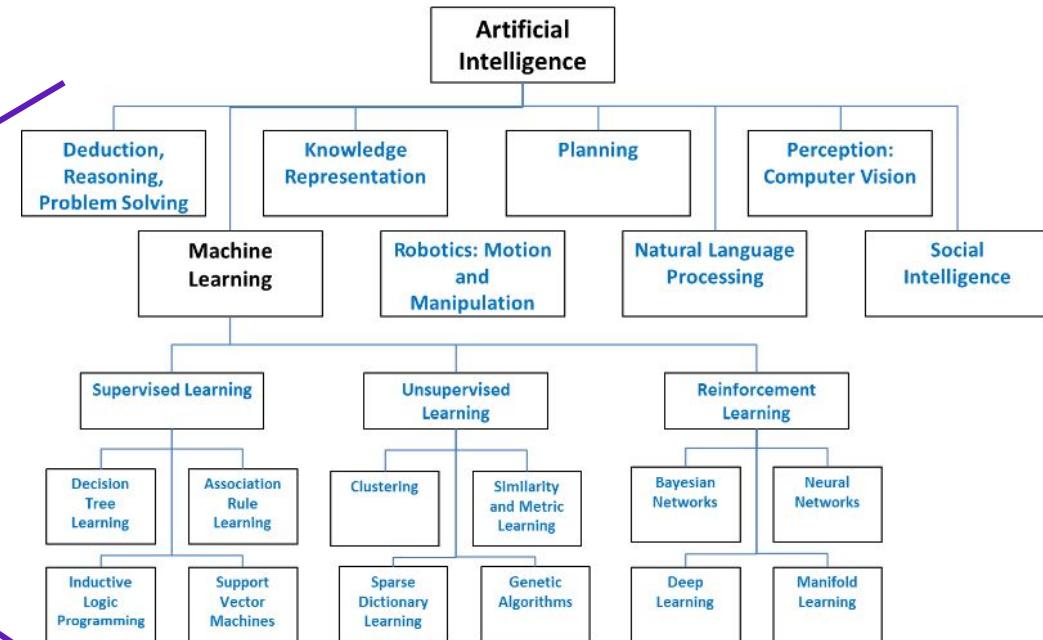
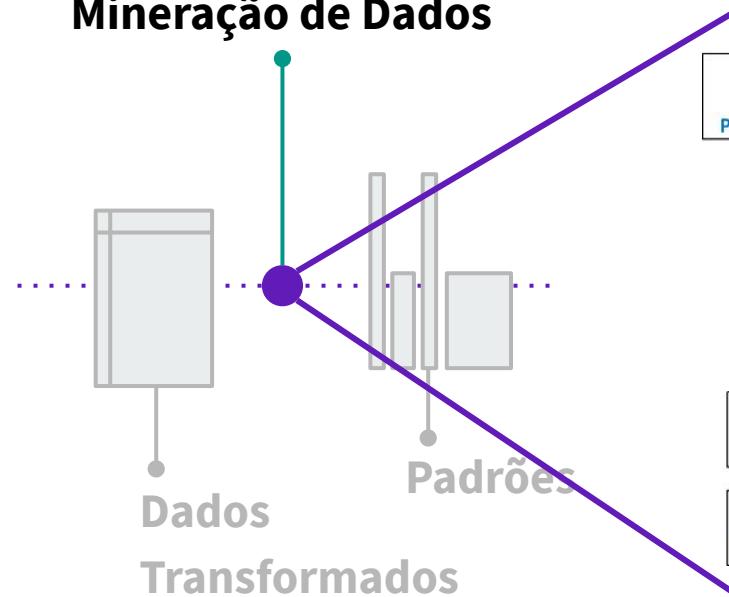
# Mineração de Dados

---



# Aprendizado de Máquina

## Mineração de Dados





## INDUSTRY 1.0

1784

Mechanization, steam power, weaving loom

## INDUSTRY 2.0

1870

Mass production, assembly line, electrical energy

## INDUSTRY 3.0

1969

Automation, computers and electronics



## INDUSTRY 4.0

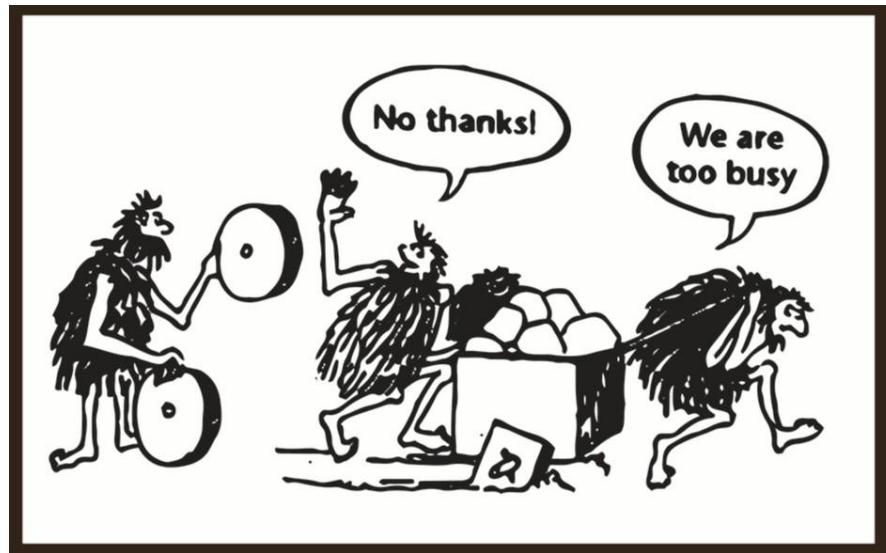
TODAY

Cyber Physical Systems, internet of things, networks artificial intelligence

# Porque usar Data Science?

---

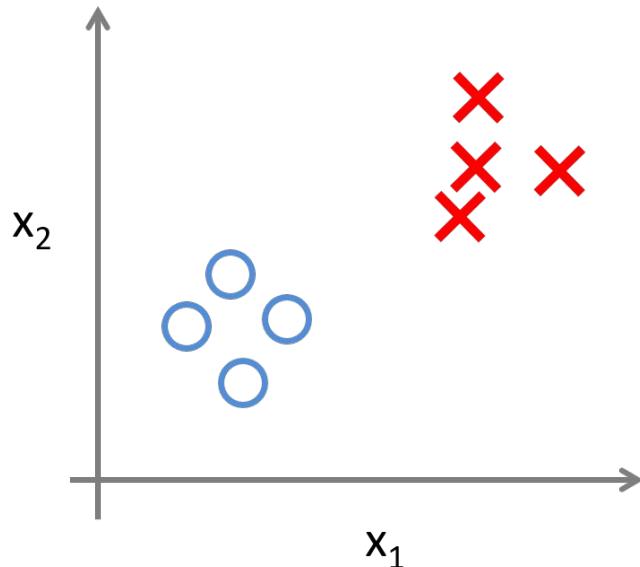
Para reduzir os custos e aumentar a eficiência da sua organização, otimizando processos.



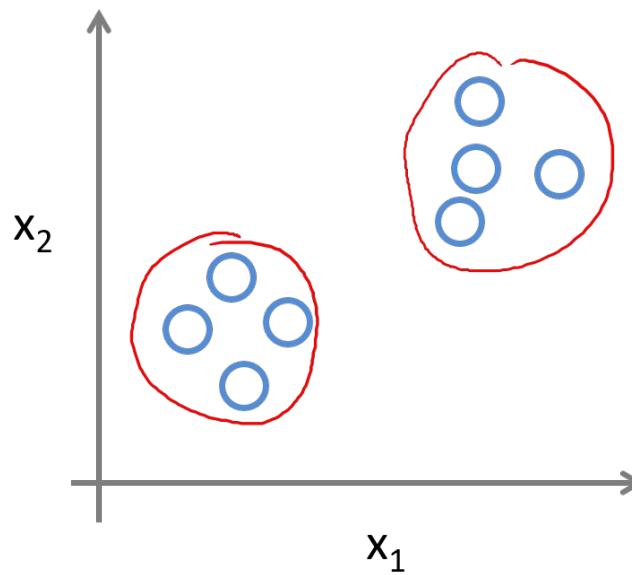
# Tipos de Aprendizagem

# Aprendizado de Máquina

Supervised Learning

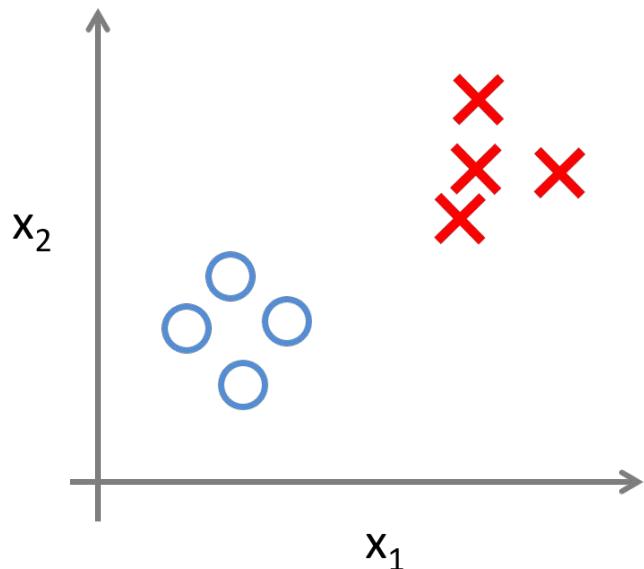


Unsupervised Learning

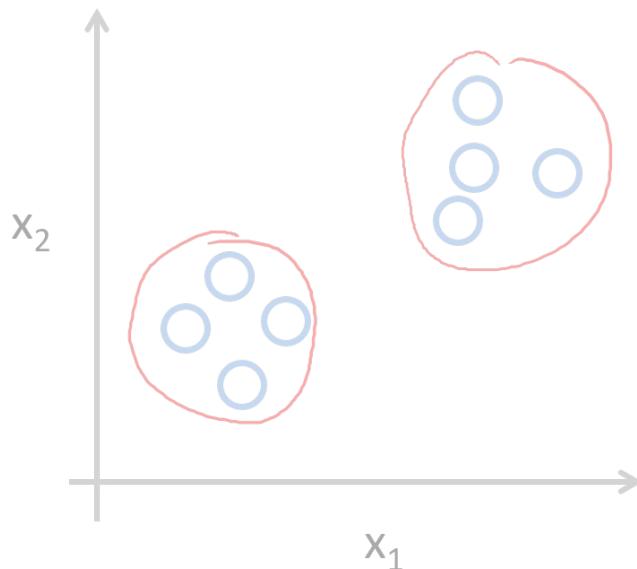


# Supervisionado

Supervised Learning



Unsupervised Learning





 alamy stock photo

JC3KYR  
[www.alamy.com](http://www.alamy.com)

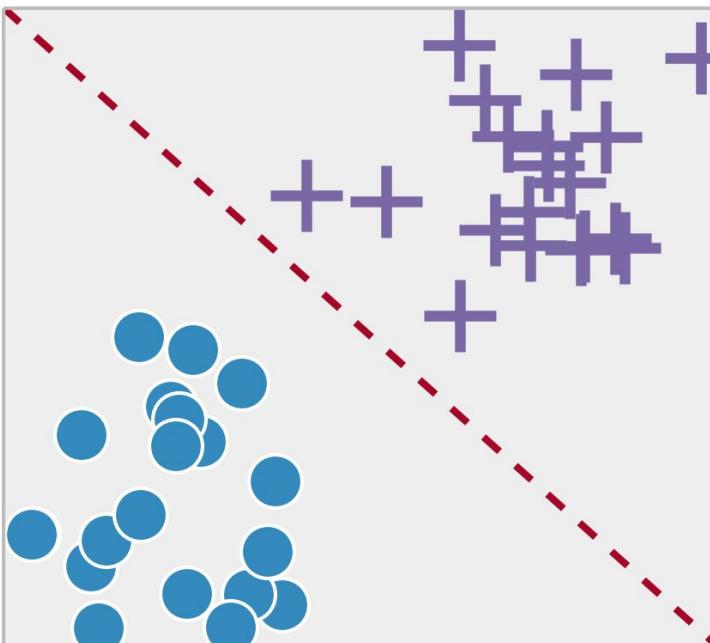


 alamy stock photo

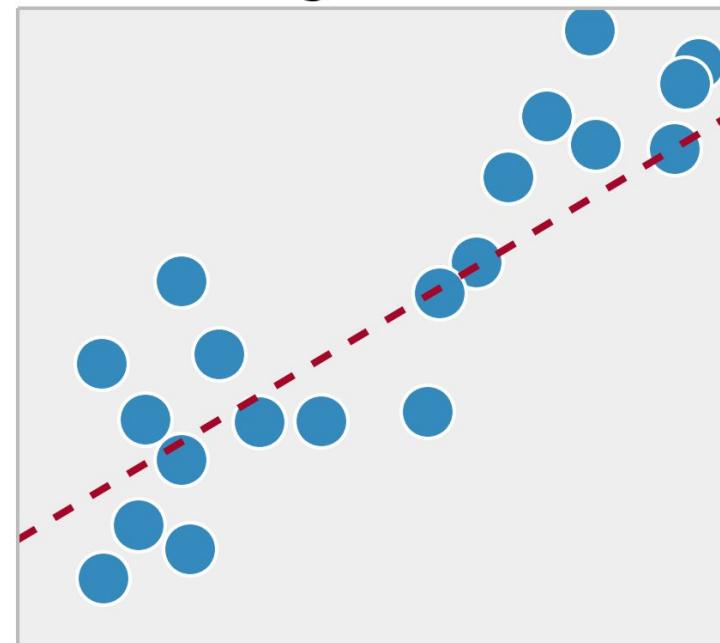
E8NGXG  
[www.alamy.com](http://www.alamy.com)

# Supervisionado

Classification



Regression



# Classificação (atributos)

---

Sintomas		Exames					Sinais			Resultado	
0	dor	1	0	1	0	0.15	8	0	10	15	...
0	febre	0	0	0	0.11	0	0	0.5	9	12	...
0	-	0	0	1	0	0.12	7	0	11	12	...
											Sim
											Não
											Sim

# Regressão (atributos)

---

Sintomas	Exames								Sinais			Resultado
0	dor	1	0	1	0	0.15	8	0	10	15	...	5
0	febre	0	0	0	0.11	0	0	0.5	9	12	...	12
0	-	0	0	1	0	0.12	7	0	11	12	...	7

Online

# Oficina Gratuita Machine Learning



AI New Ventures

31/10 - 14h às 17h  
Transmissão via [Youtube](#).

Intervalo de 15 min ...

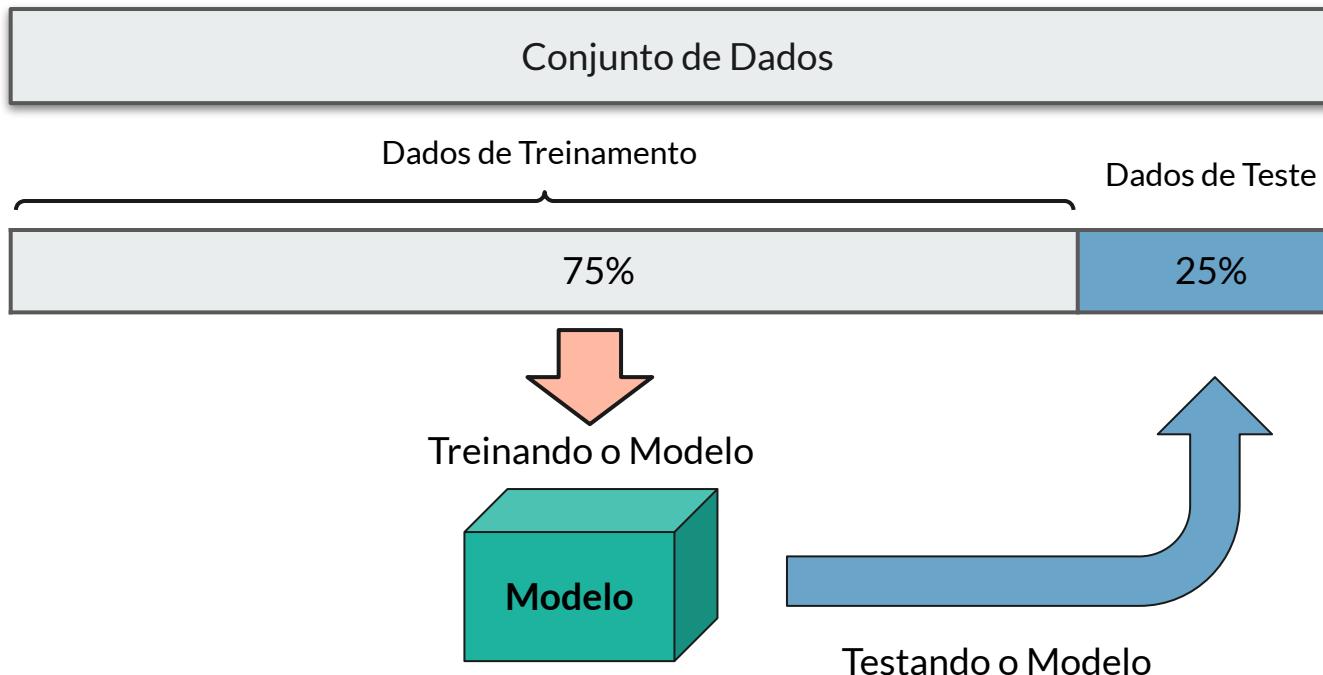
Instituição beneficiada



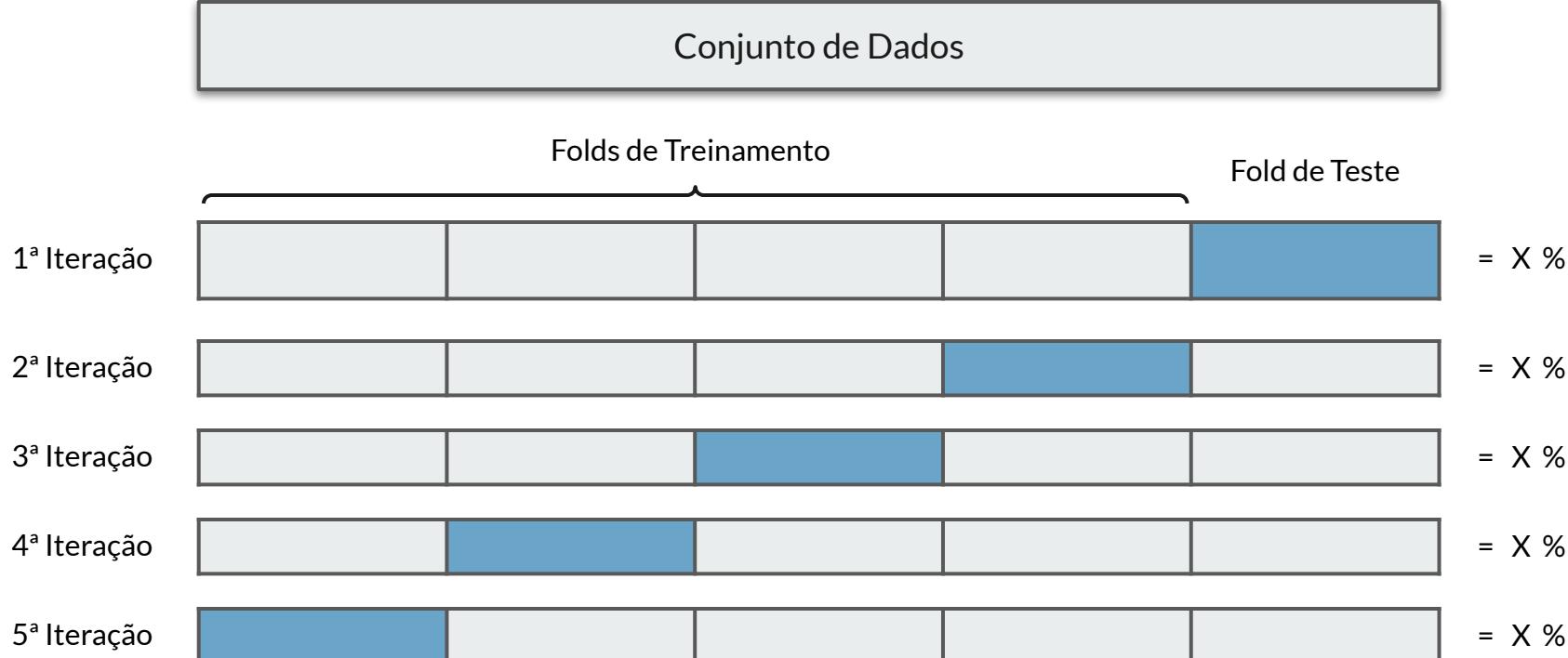
*Instituto  
do Câncer  
Infantil*

# Avaliação dos Modelos

# Treino e Teste

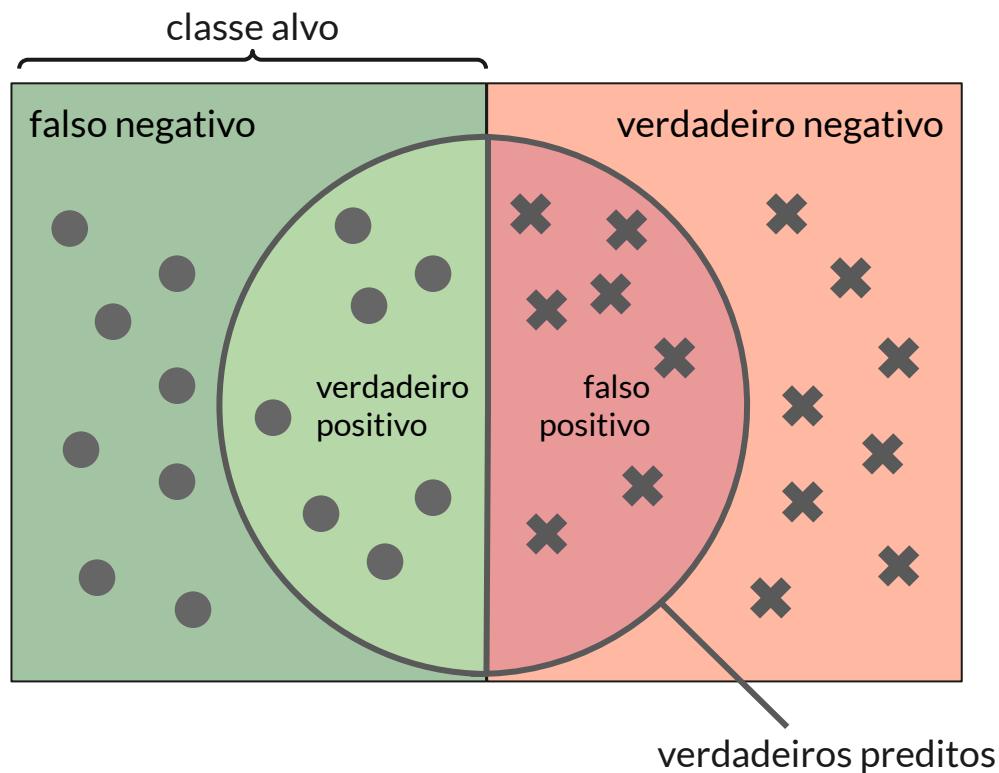


# Validação Cruzada (5 Folds)



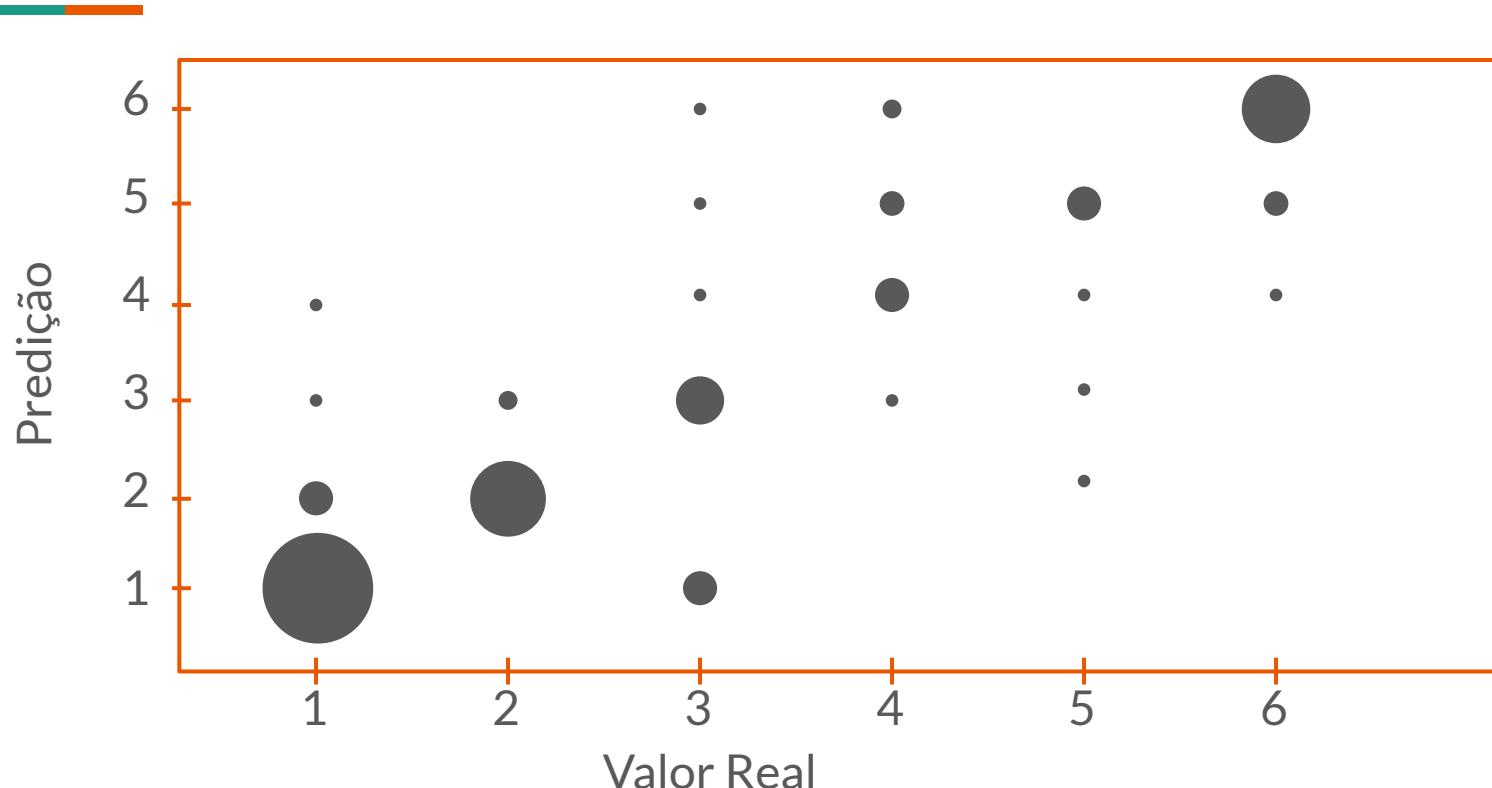
# Acurácia, Precisão e Abrangência

## Classificação

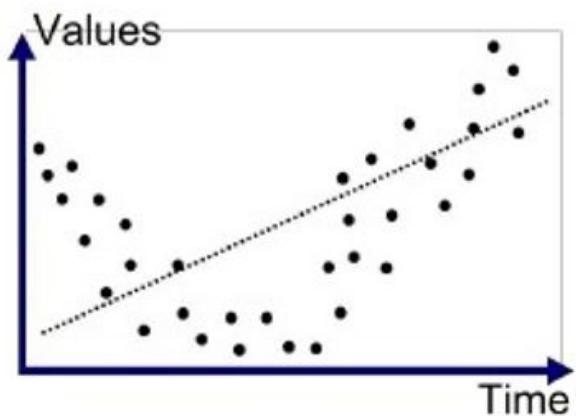


$$\text{Acurácia} = \frac{\text{verdadeiro positivo} + \text{verdadeiro negativo}}{\text{falso negativo} + \text{verdadeiro positivo} + \text{falso positivo} + \text{verdadeiro negativo}}$$
$$\text{Precisão} = \frac{\text{verdadeiro positivo}}{\text{verdadeiro positivo} + \text{falso positivo}}$$
$$\text{Abrangência} = \frac{\text{verdadeiro positivo}}{\text{verdadeiro positivo} + \text{falso negativo}}$$

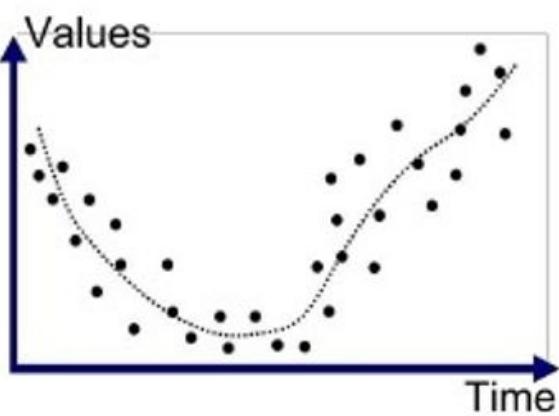
# Erro Médio Absoluto Regressão



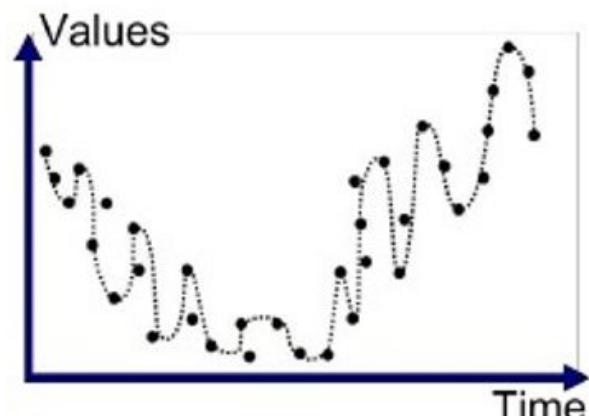
# Overfitting na Regressão



Underfitted

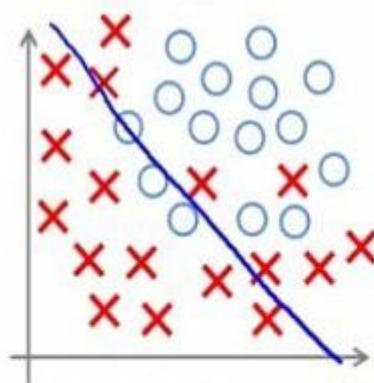


Good Fit/R robust

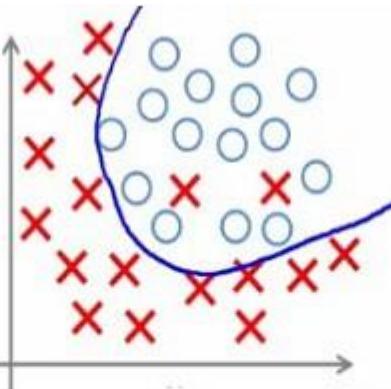


Overfitted

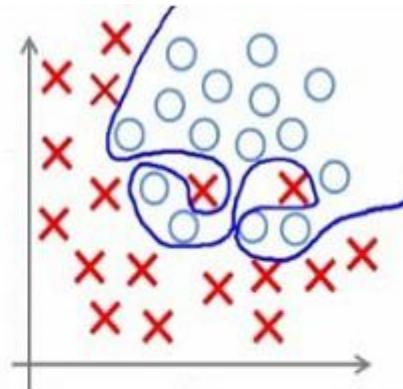
# Overfitting na Classificação



Under-fitting



Appropriate-fitting



Over-fitting

(too simple to explain the variance)

(overfitting -- too good to be true)

# Ética na Inteligência Artificial

# Explainable A.I.

Today



Training  
Data



Learning  
Process



Learned  
Function



Output



User with  
a Task

---

Tomorrow



Training  
Data



New Learning  
Process



Explainable  
Model



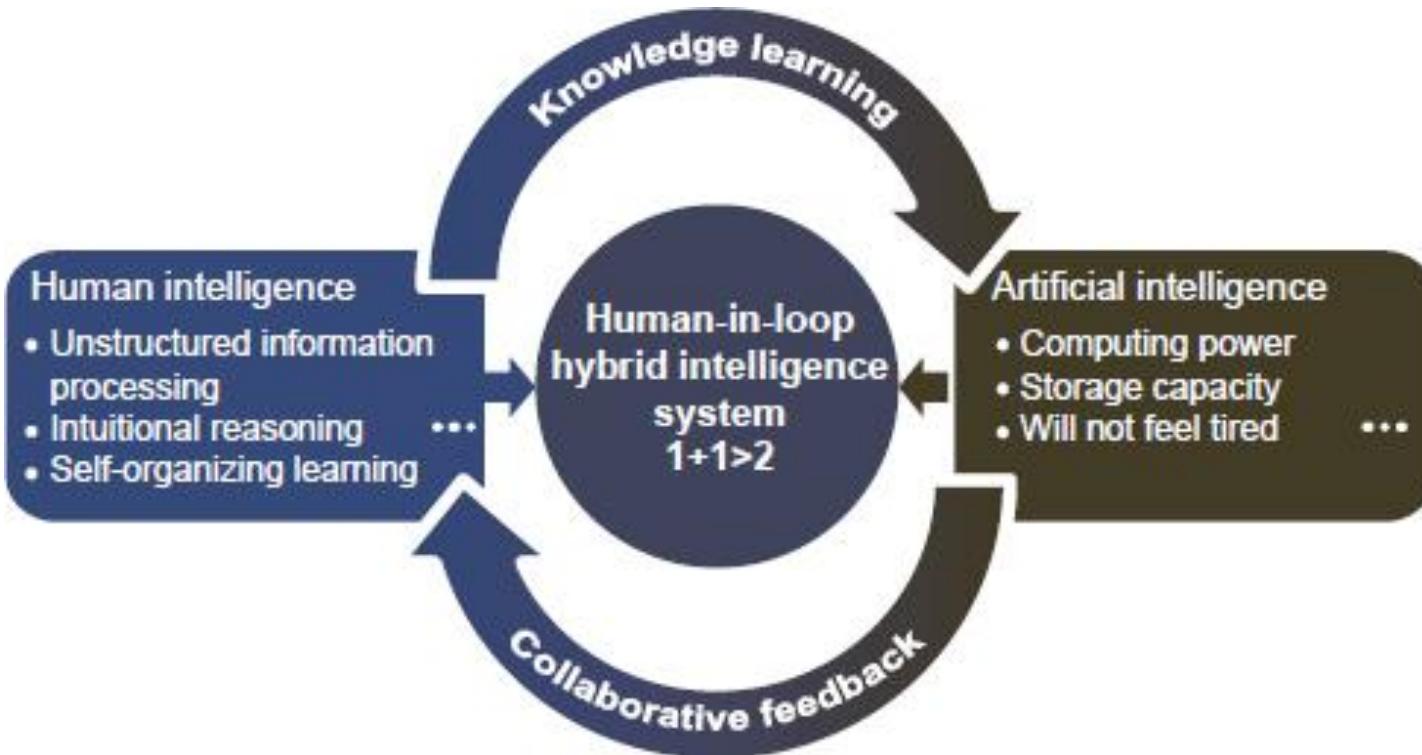
Explainable Interface



User with  
a Task

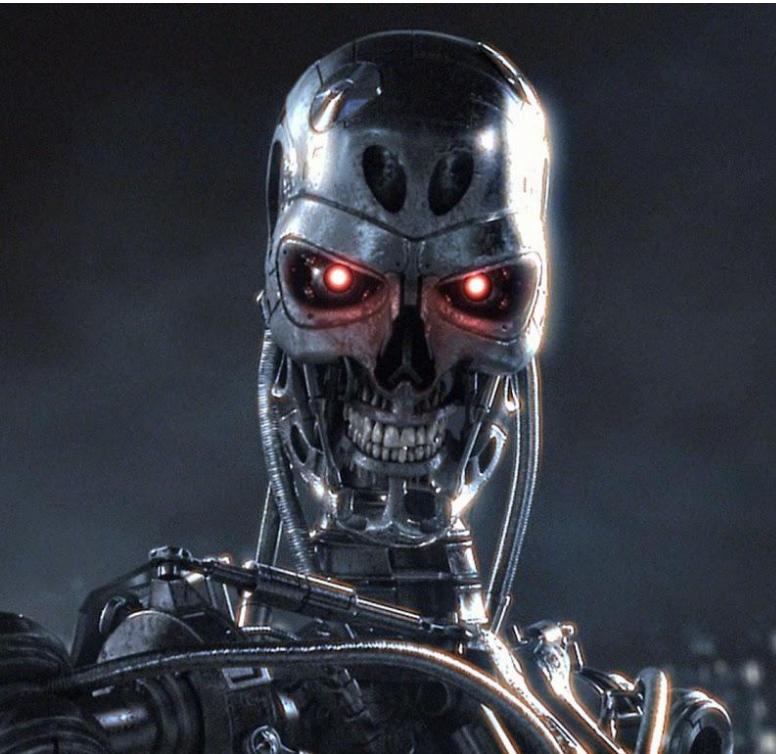
# Human in the Loop

---



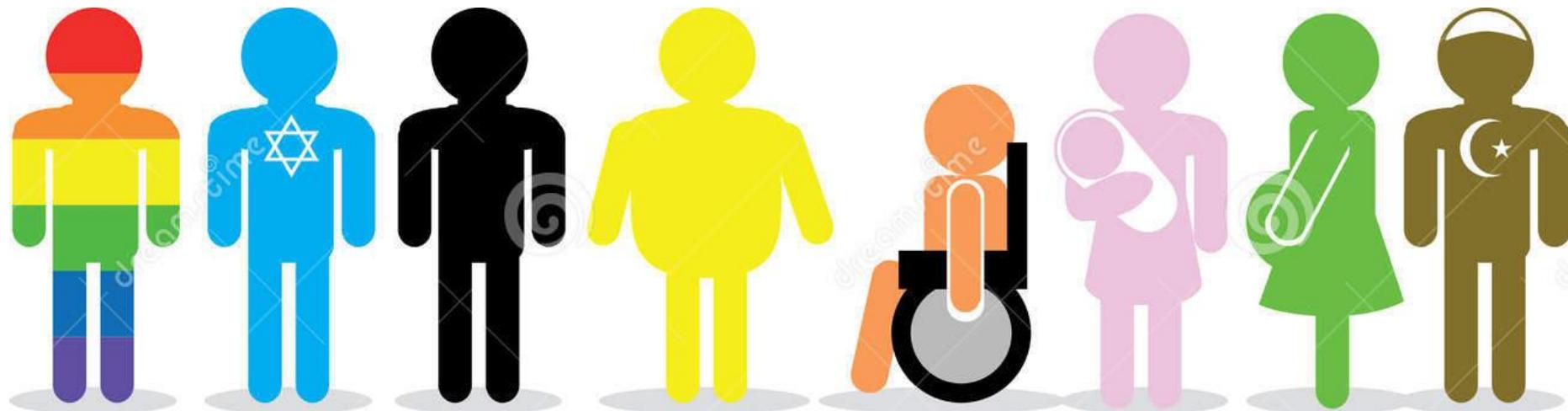
# I.A. Genérica

---



# Ética em I.A. Representando Minorias

---



Manifesto de Asilomar: <https://futureoflife.org/ai-principles/>

# Obrigado!

Henrique Dias

Cientista de Dados

[henrique@noharm.ai](mailto:henrique@noharm.ai)

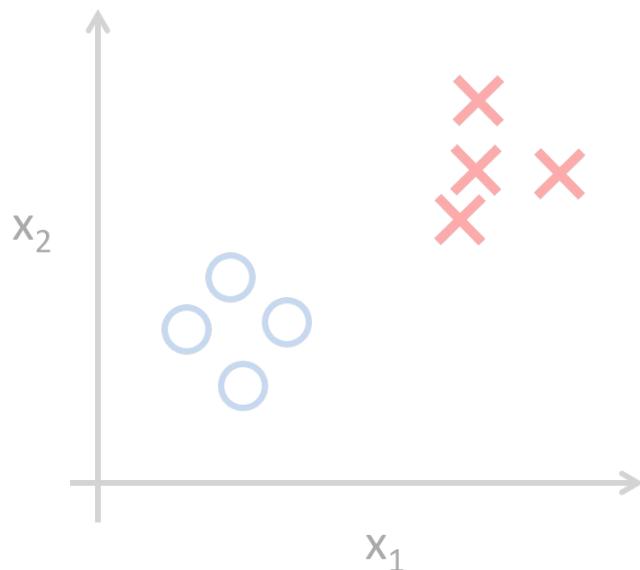
[PDF dos Slides](#)



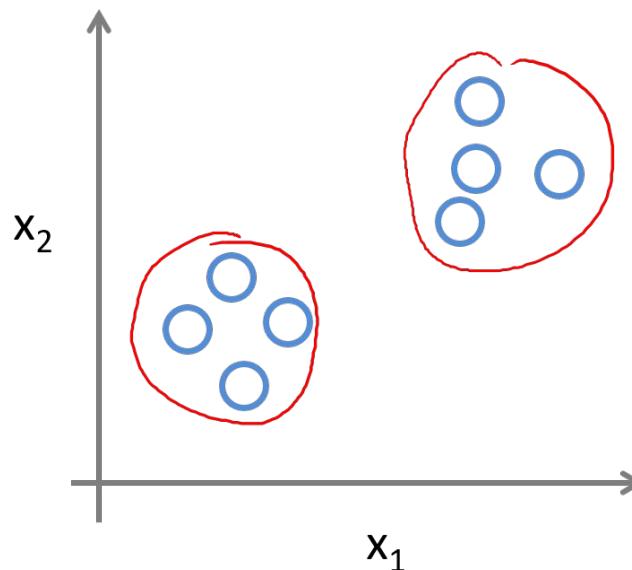
# Aprendizado Não-supervisionado

# Não-Supervisionado

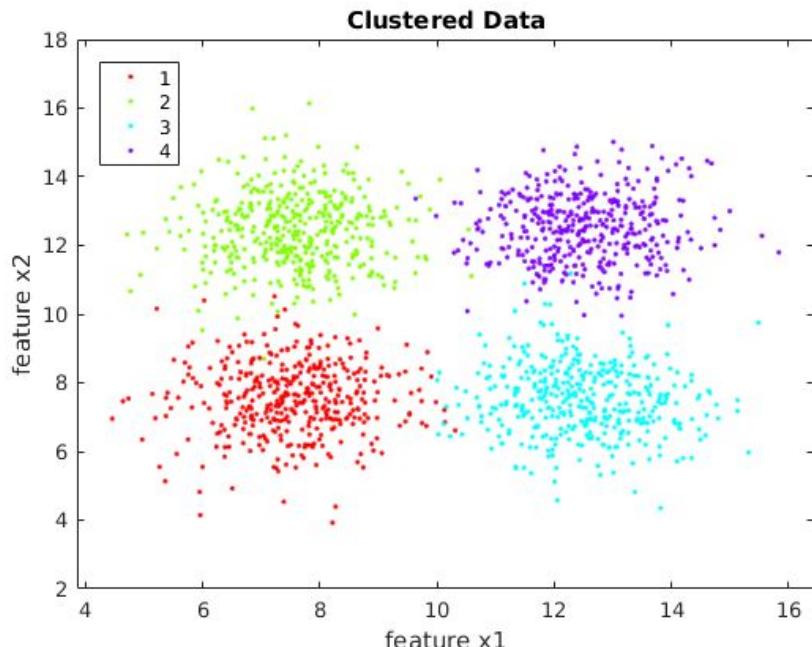
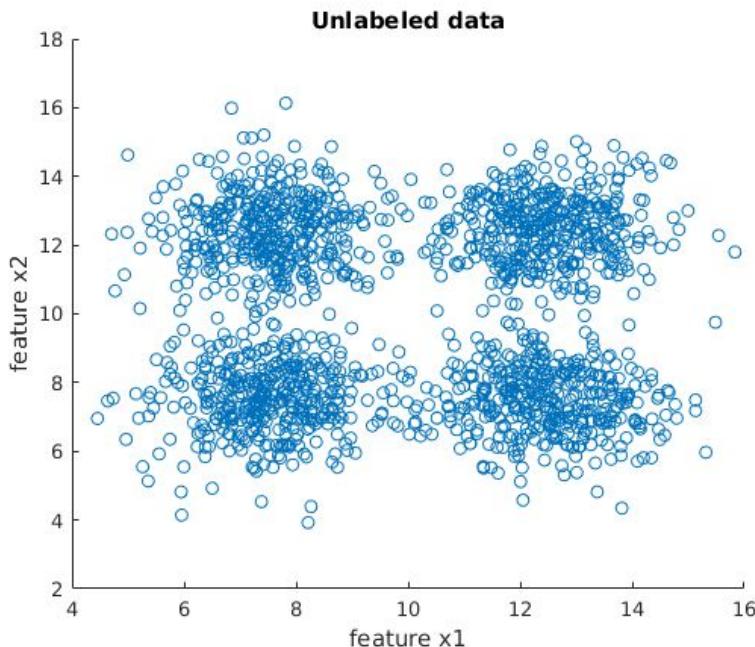
Supervised Learning



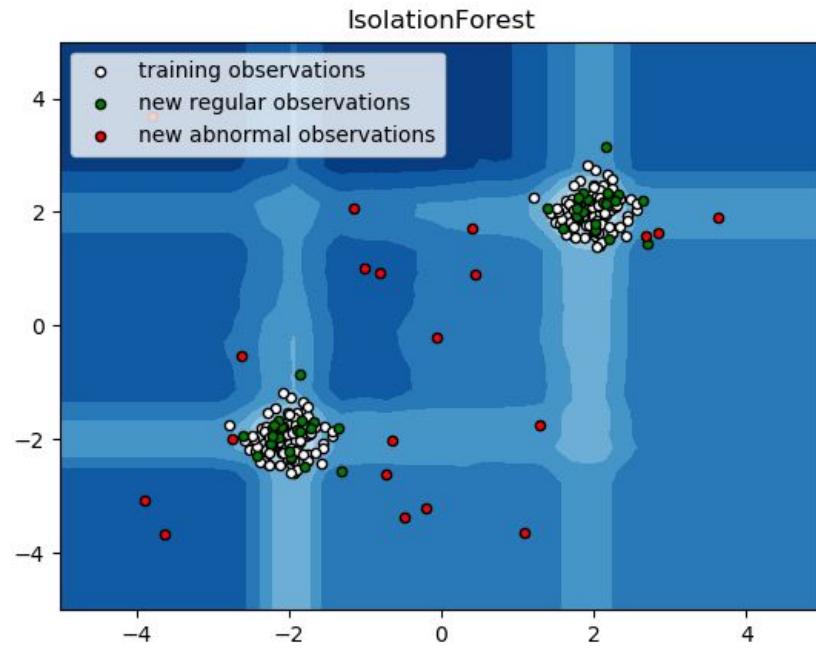
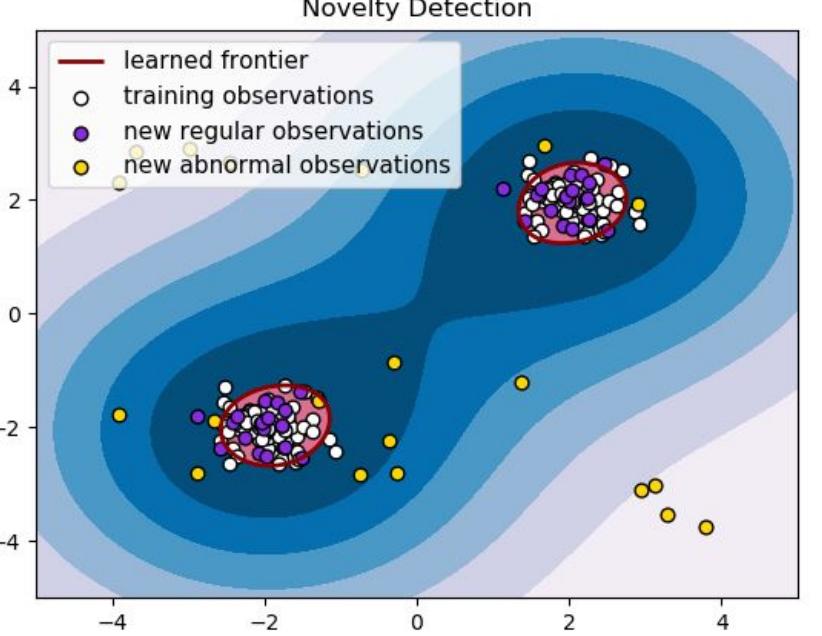
Unsupervised Learning

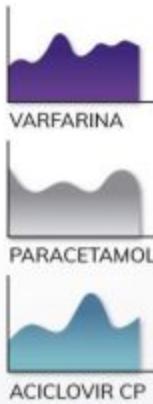


# Não-Supervisionado **Agrupamento**



# Não-Supervisionado Detecção de Anomalia





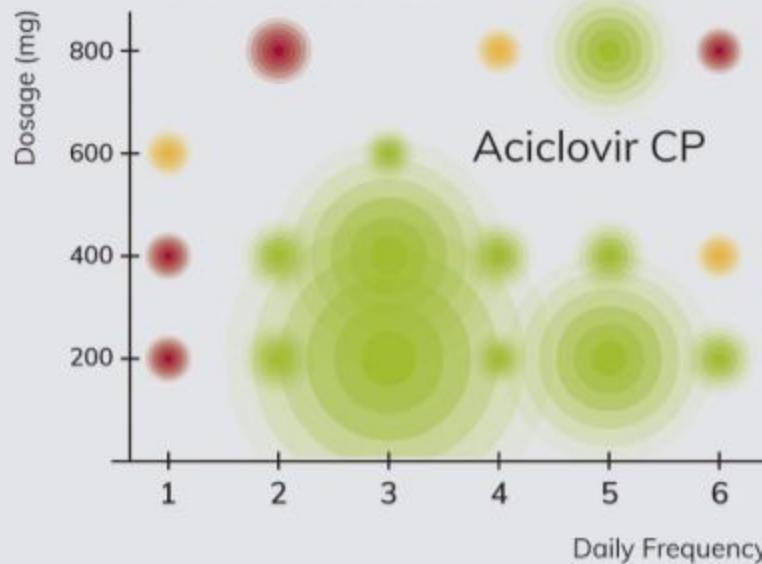
## Medication's Distribution



## Prescriptions



# Density Distance Centrality Outlier



## Smart Prescription Screening

### PATIENTS PRESCRIPTION

John Smith - 02/15/2018

Alan Moore - 02/02/2018

Caterina Jonhson - 02/10/2018

Helena Troy - 02/11/2018

Henry Angels - 02/13/2018

### MEDICATIONS

CLOPIDOGREL 75mg CP

PARACETAMOL 500 mg CP

OMEPRAZOL 20 mg

DIPIRONA 1 g/2 ml SOL INJ

SINVASTATINA 20mg CP

Open-Source: <https://github.com/nlp-pucrs/ddc-api>

# Não-Supervisionado Regras de Associação



Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B \& C \Rightarrow D$	1/5	1/3	5/9

$$Supprt = \frac{Frequency(X, Y)}{N}$$

$$\rightarrow Confidence = \frac{Frequency(X, Y)}{Frequency(X)}$$

$$Lift = \frac{Support}{Support(X) \times Support(Y)}$$

# Não-Supervisionado **Redução de Dimensionalidade**

First centered Olivetti faces



genfaces - PCA using randomized SVD - Train time 0.1



PCA, NMF, SVD, LDA, LSA, T-SNE

# Não-Supervisionado Word Weight



$$M_{ij} = \frac{\# \text{ of word } i}{\# \text{ of word } i \text{ in doc } j}$$

0.5	-1.4	-1.4	-0.5	-1.4
-1.7	0.5	0.5	0.5	0.5
-0.5	0.5	0.5	-1.8	-0.7
0.5	0.5	0.5	0.5	0.5
-1.1	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5
	0.5	0.5	0.5	0.5

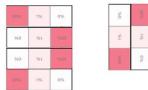
document-word

$$SVD(M_{ij})$$



document-word

LSI + Dirichlet Prior



document-word

$$M_{ij} = \frac{\#(i, j) / n_{pairs}}{\#(i) / n_{words} \#(j) / n_{words}}$$

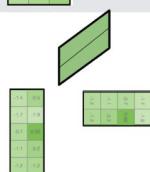
$$SVD(M_{ij})$$



word-word

$$M_{ijk} = \frac{\#(i, j, k) / n_{triplets}}{\#(i) / n_{words} \#(j) / n_{words} \#(k) / n_{words}}$$

$$SVD(M_{ijk})$$

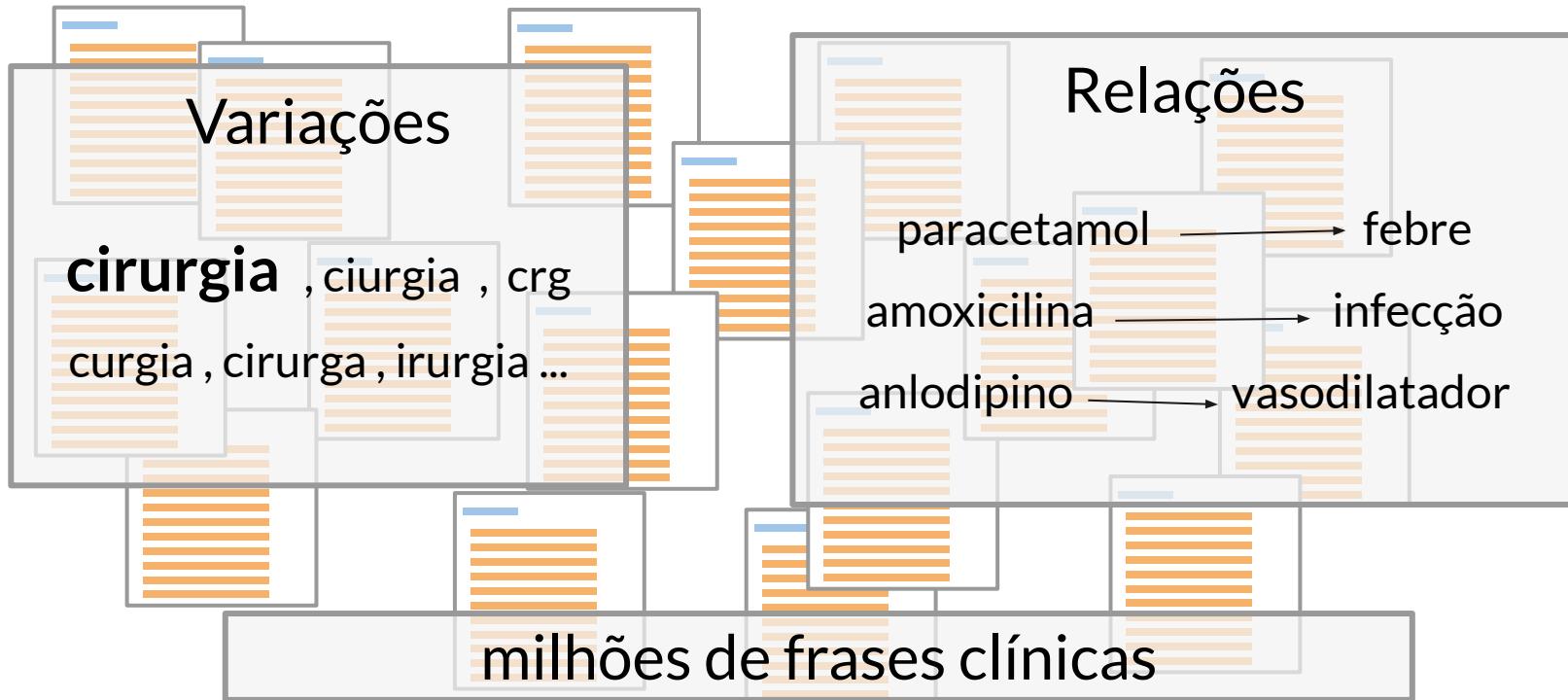


document-word-word

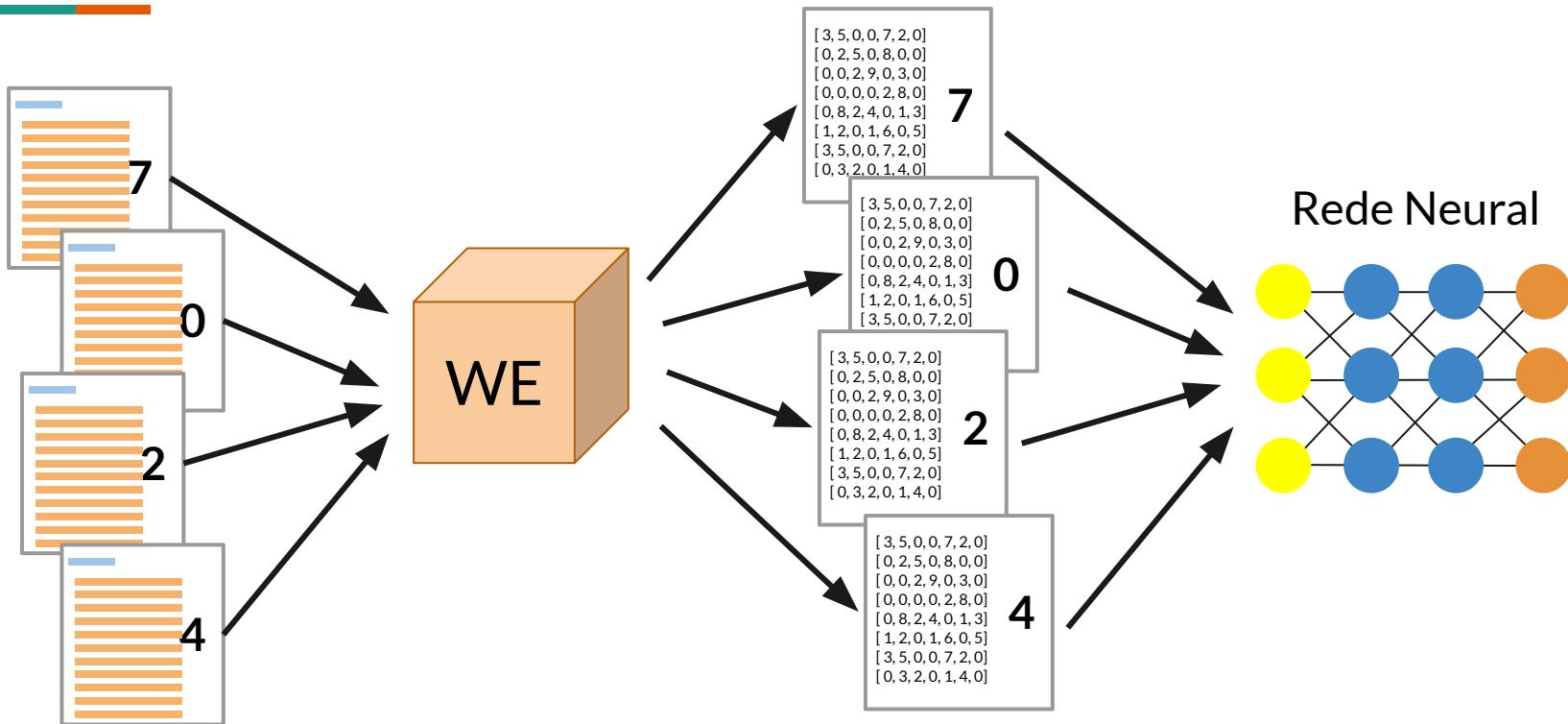
# Não-Supervisionado Word Embeddings



# Não-Supervisionado Word Embeddings



# Não-Supervisionado Word Embeddings

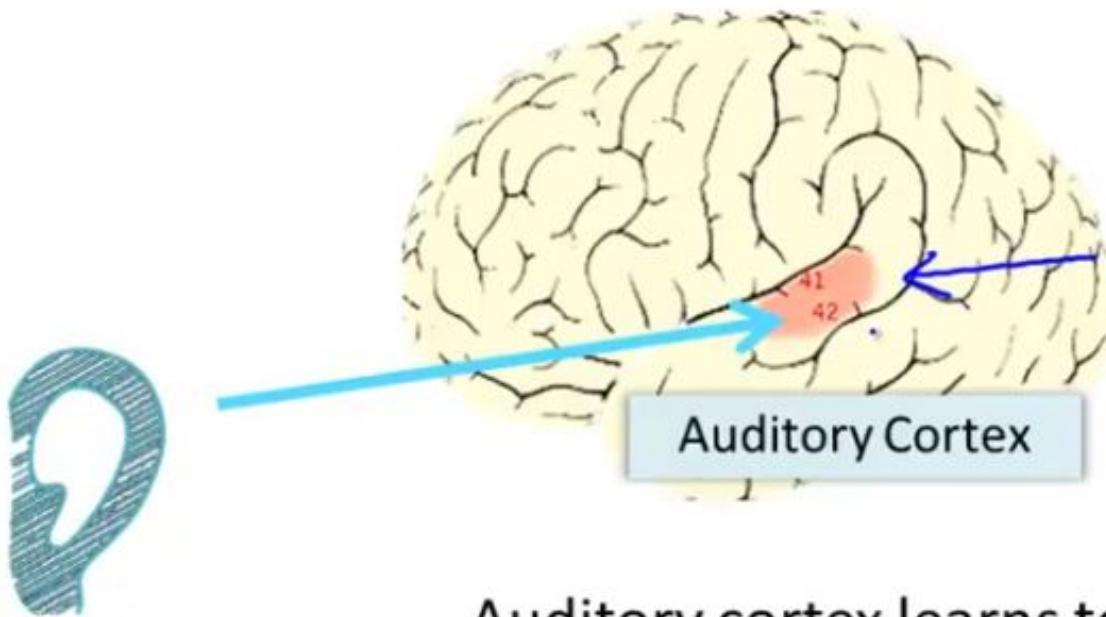


# Redes Neurais



# A hipótese do algoritmo de aprendizado único

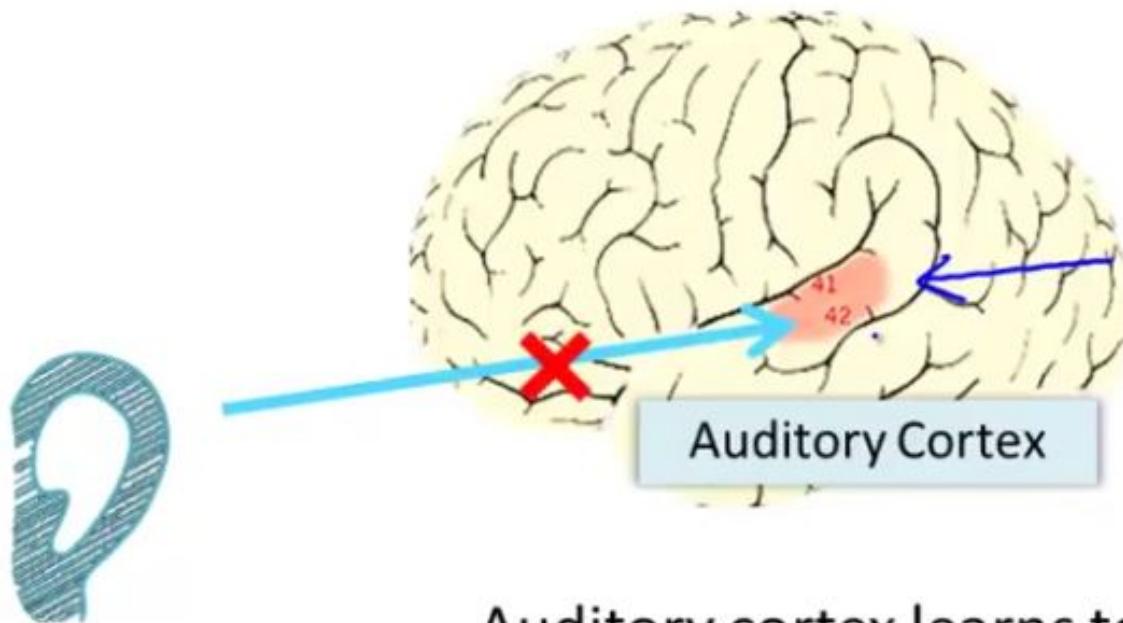
---



Auditory cortex learns to see

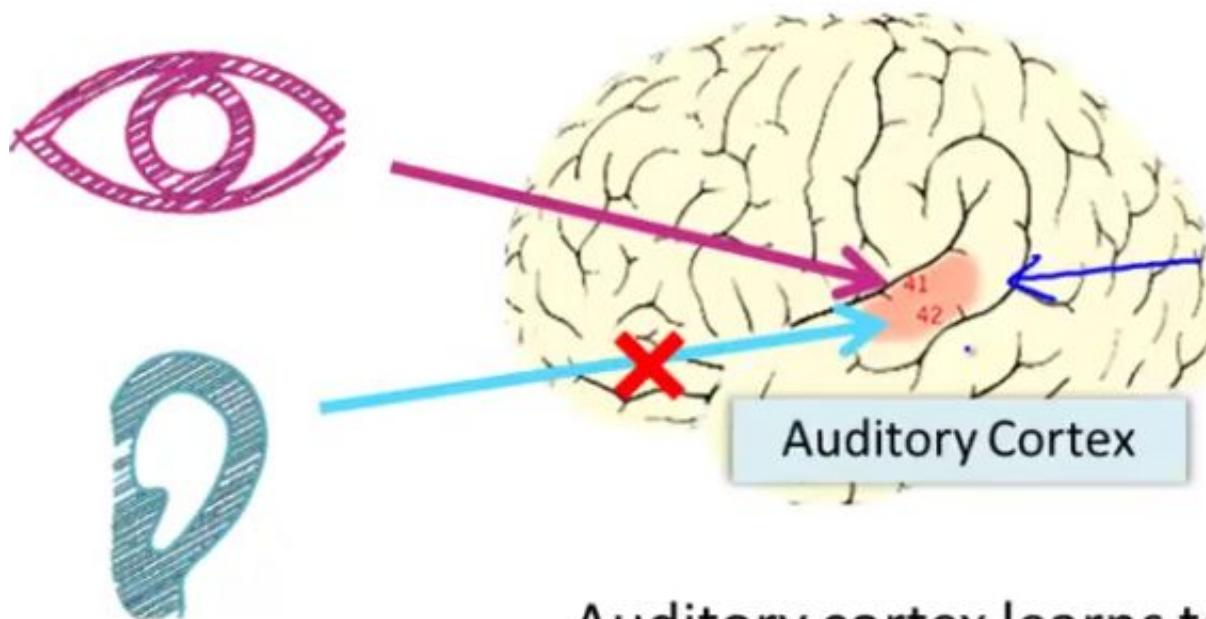
# A hipótese do algoritmo de aprendizado único

---



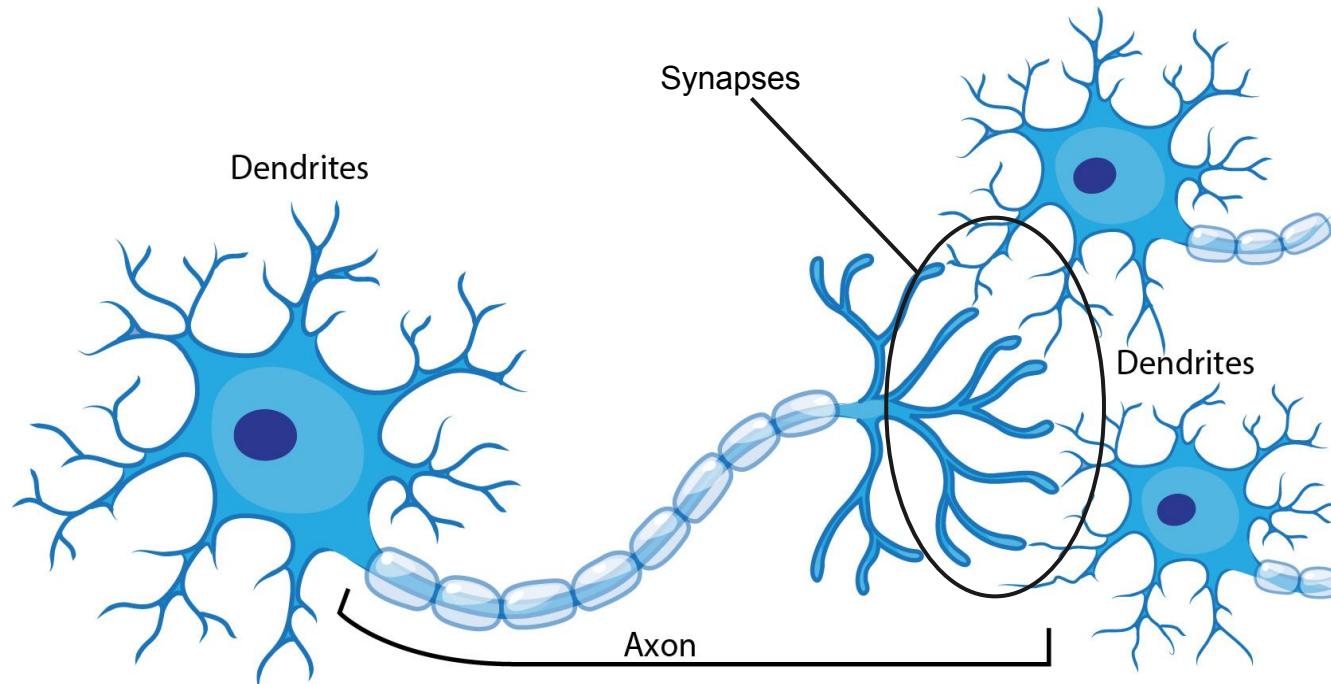
Auditory cortex learns to see

# A hipótese do algoritmo de aprendizado único

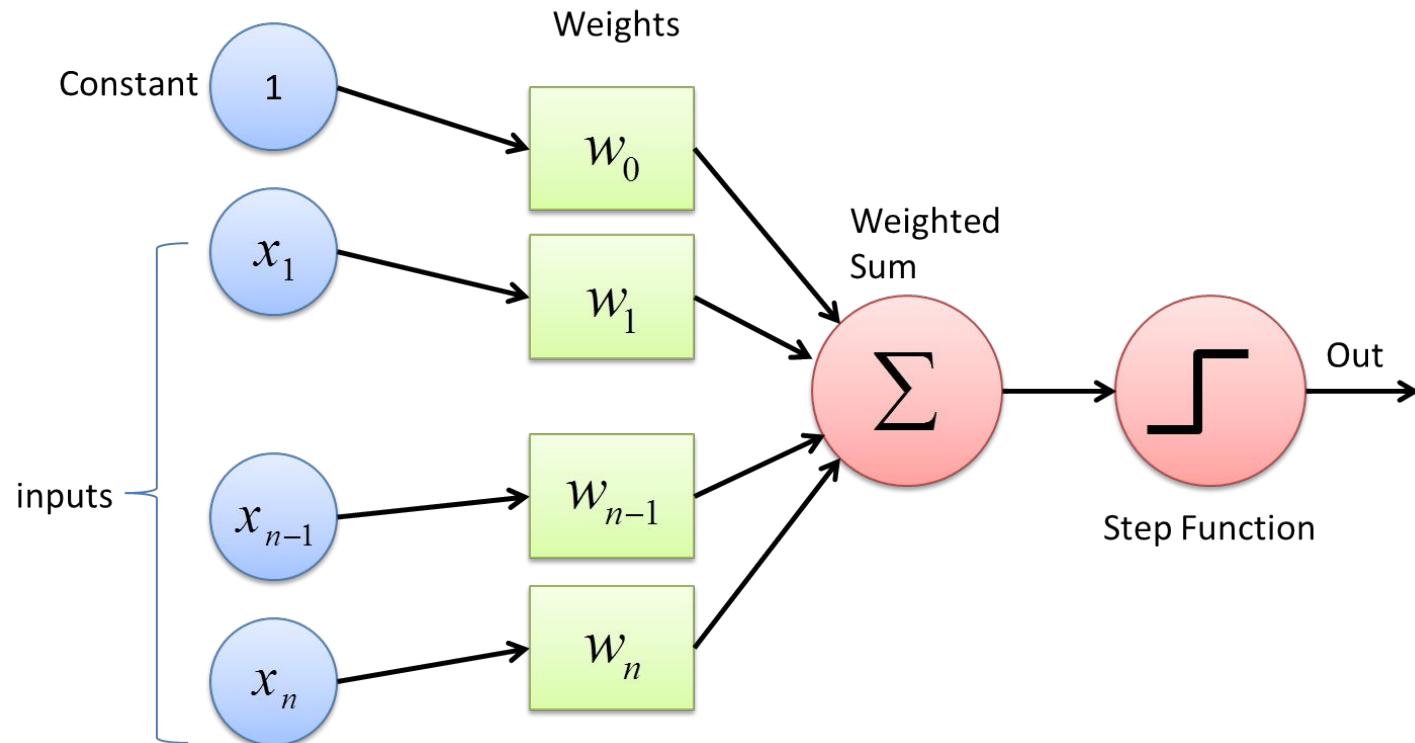


Auditory cortex learns to see

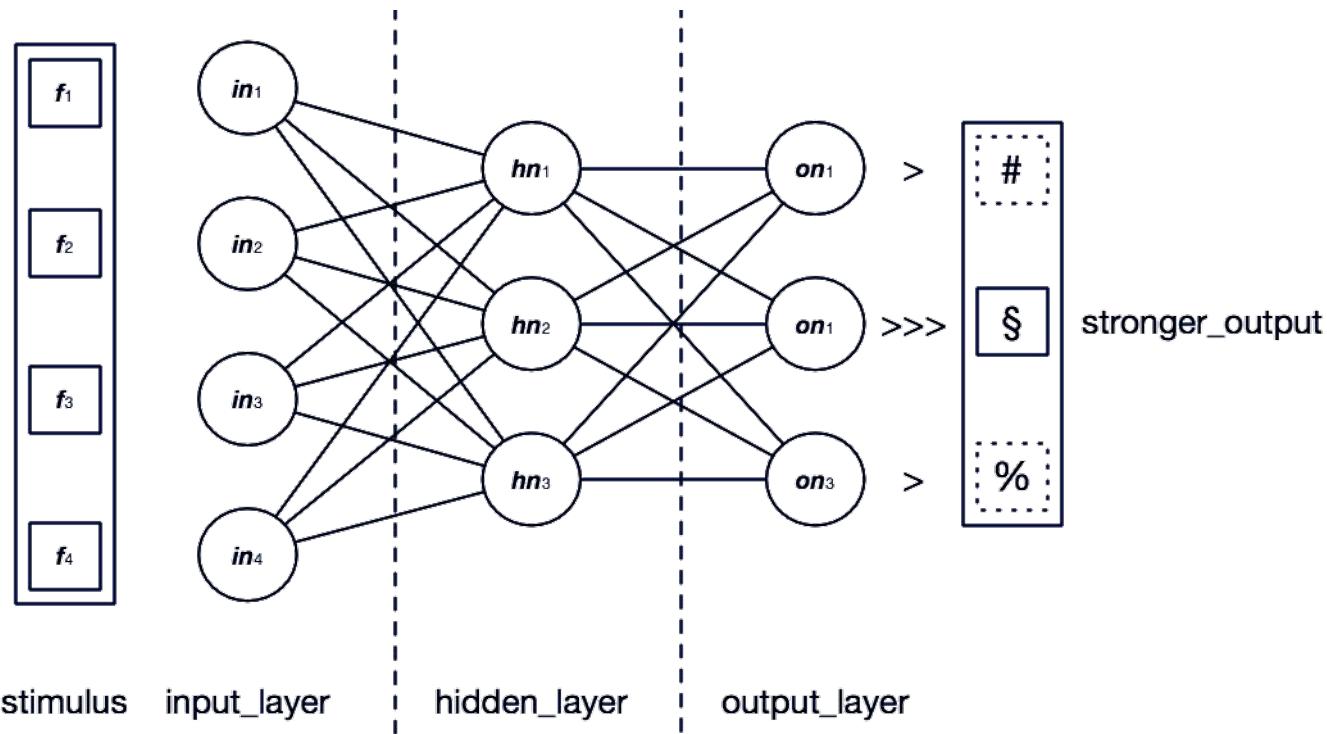
# Neurônio (perceptron)



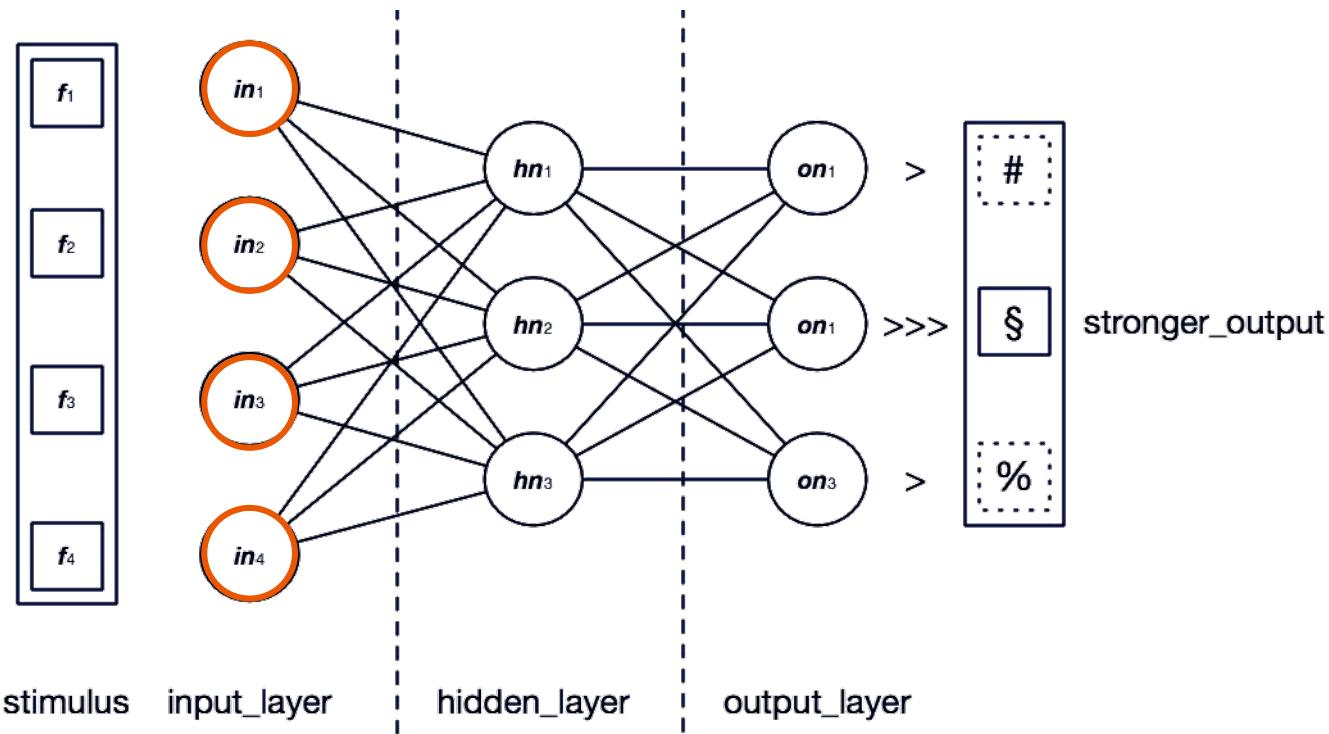
# Perceptron (neurônio)



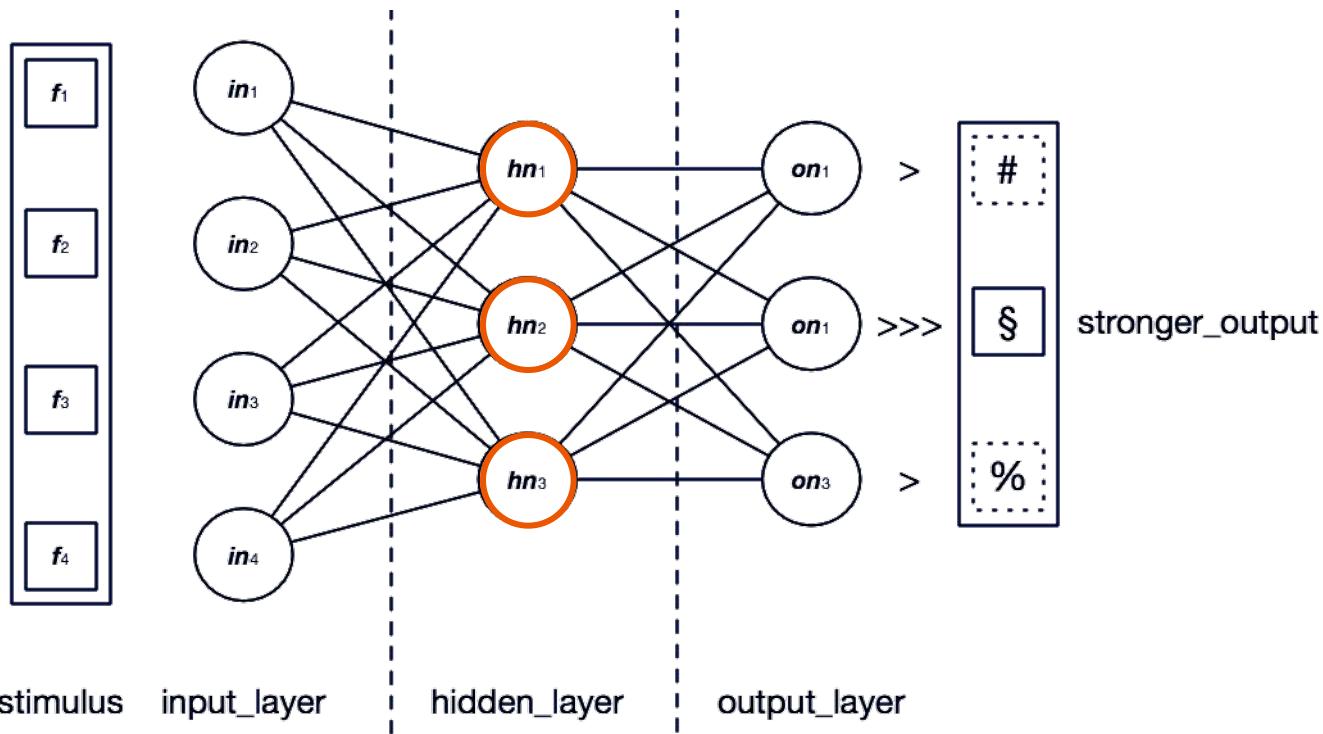
# Multi Layer Perceptron



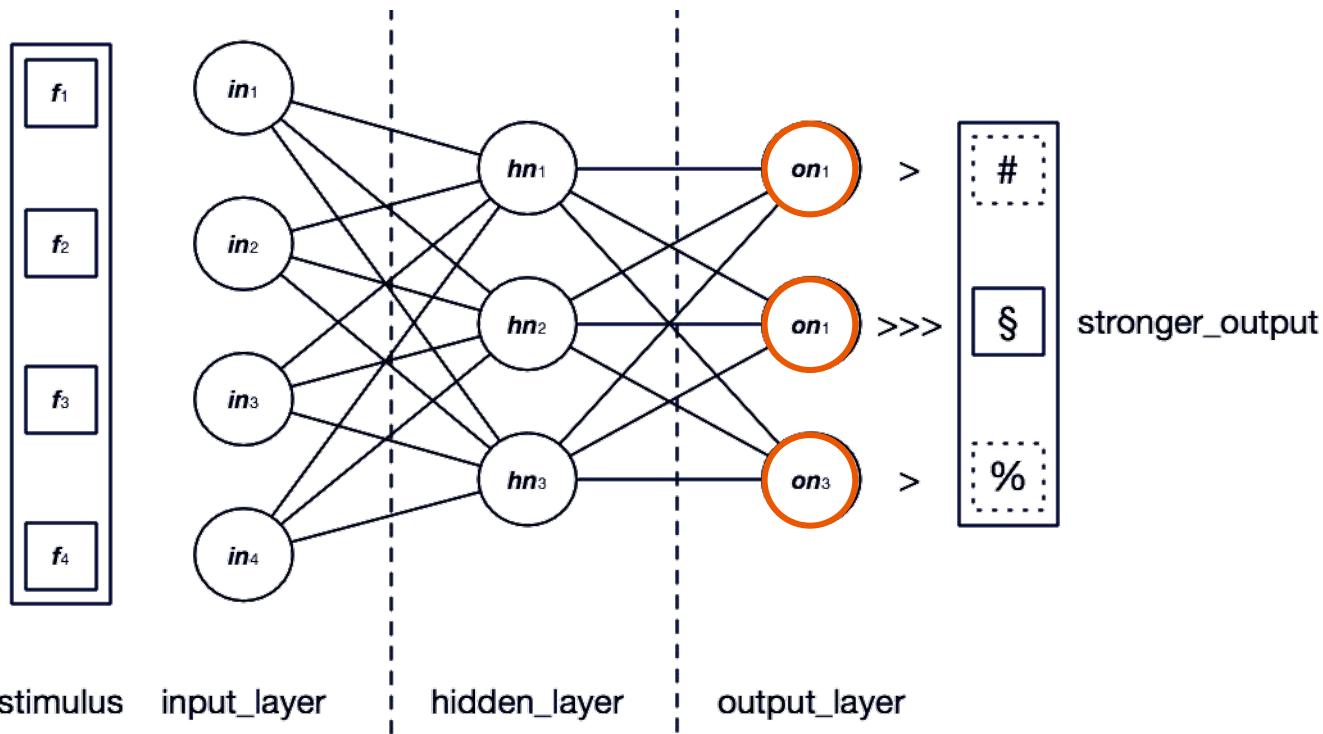
# Multi Layer Perceptron



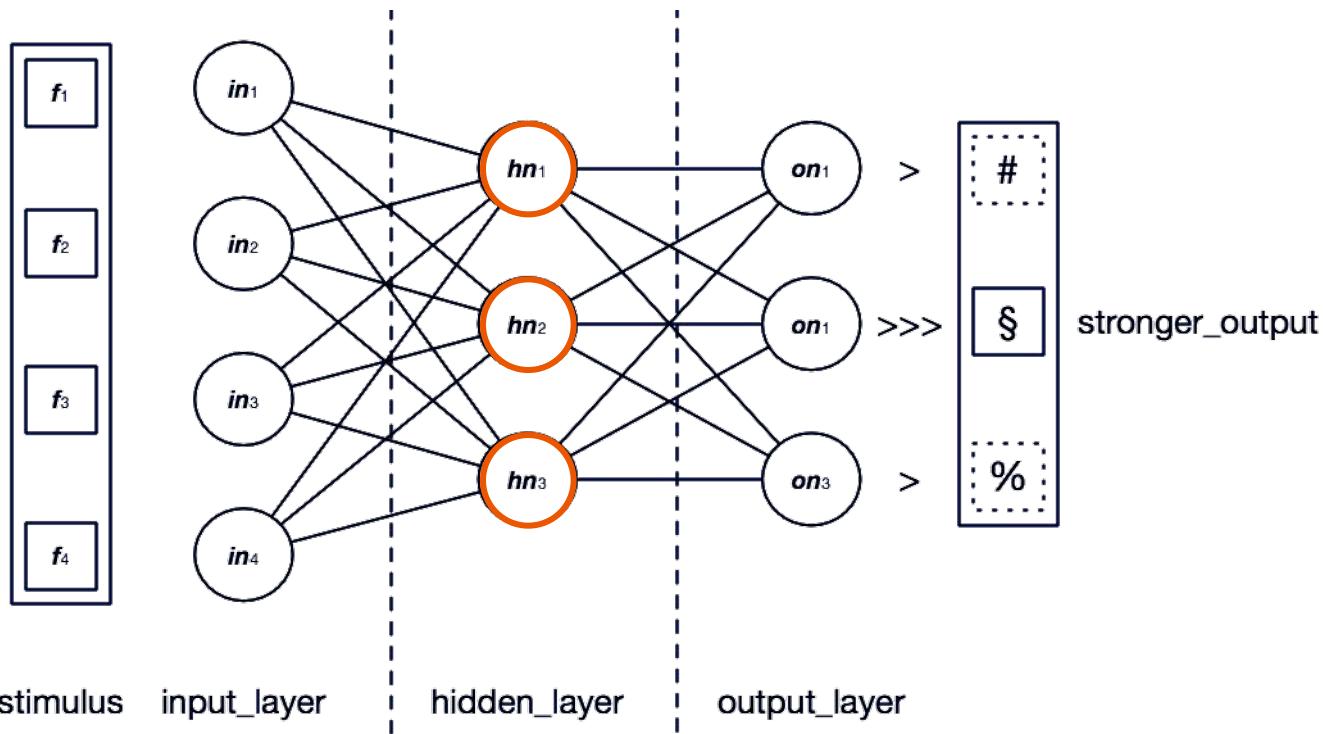
# Multi Layer Perceptron



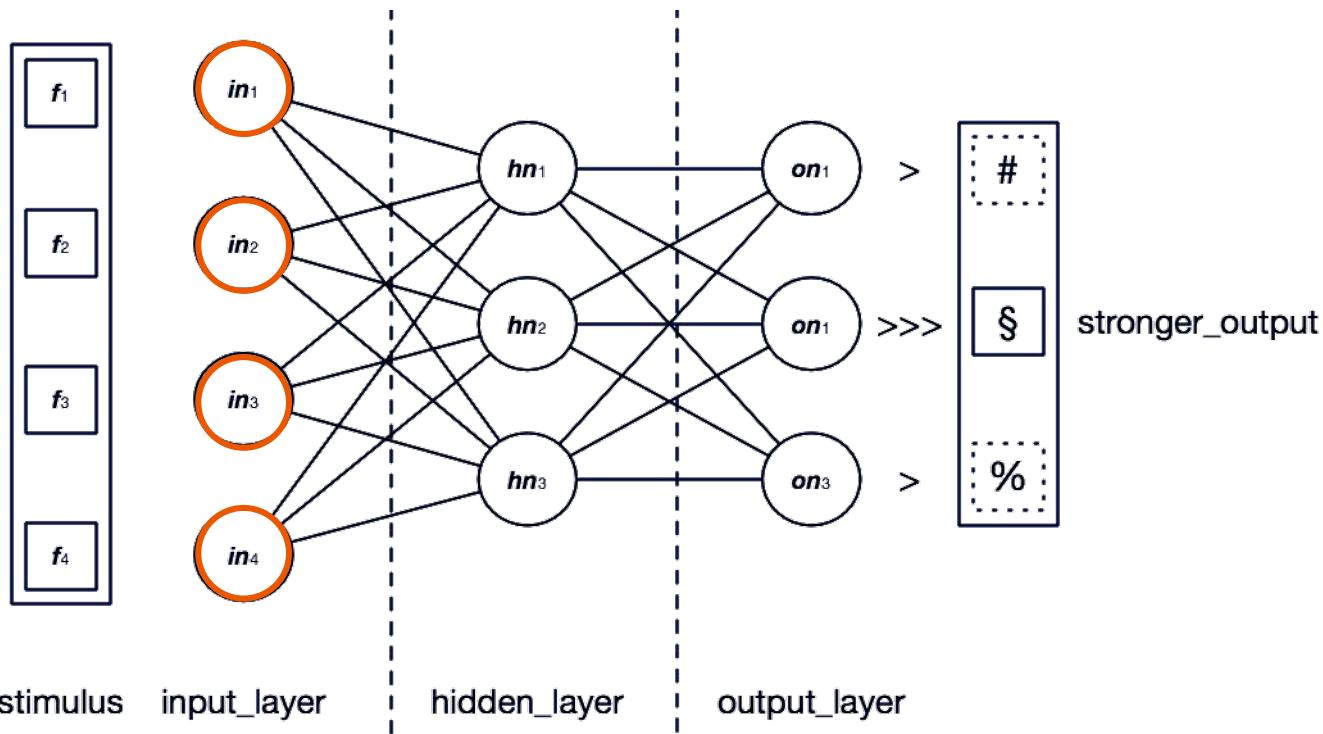
# Multi Layer Perceptron



# Multi Layer Perceptron



# Multi Layer Perceptron



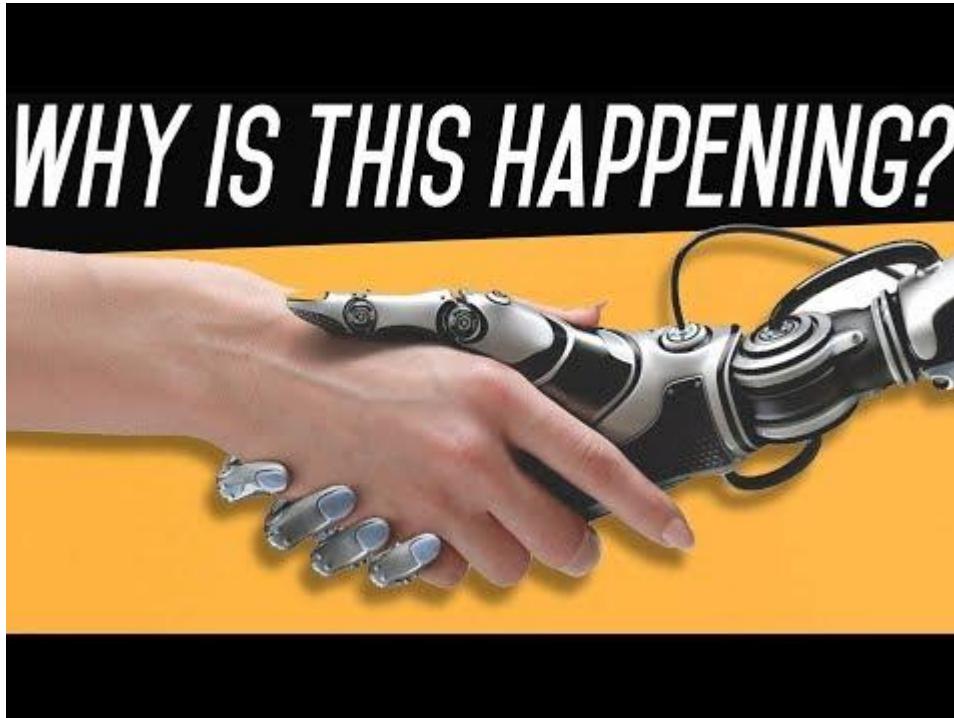
# Entendendo Redes Neurais

---

<https://playground.tensorflow.org>

# Do que as RNs são capazes?

---

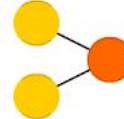


# Neural Networks

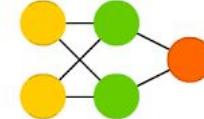
-  Backfed Input Cell
-  Input Cell
-  Noisy Input Cell
-  Hidden Cell
-  Probabilistic Hidden Cell
-  Spiking Hidden Cell
-  Output Cell
-  Match Input Output Cell
-  Recurrent Cell
-  Memory Cell
-  Different Memory Cell
-  Kernel
-  Convolution or Pool

©2016 Fjodor van Veen - [asimovinstitute.org](http://asimovinstitute.org)

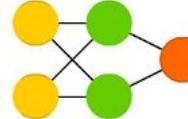
Perceptron (P)



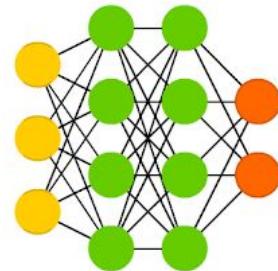
Feed Forward (FF)



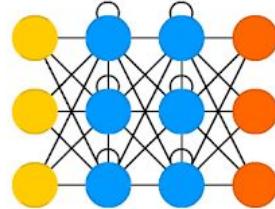
Radial Basis Network (RBF)



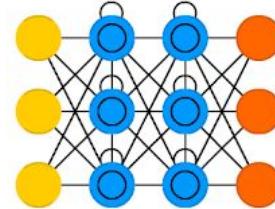
Deep Feed Forward (DFF)



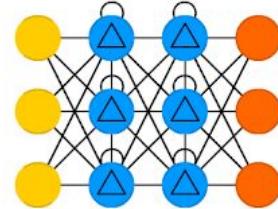
Recurrent Neural Network (RNN)



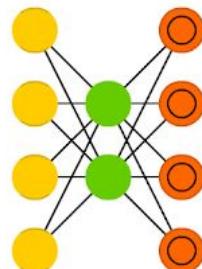
Long / Short Term Memory (LSTM)



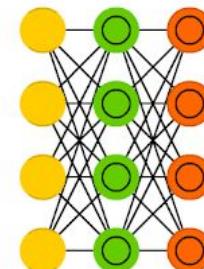
Gated Recurrent Unit (GRU)



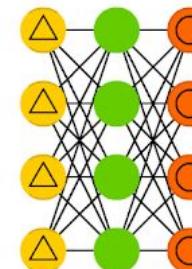
Auto Encoder (AE)



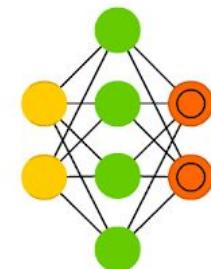
Variational AE (VAE)



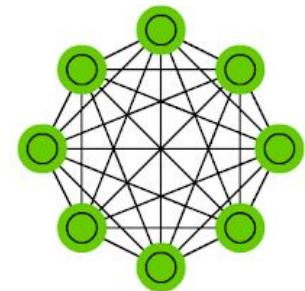
Denoising AE (DAE)



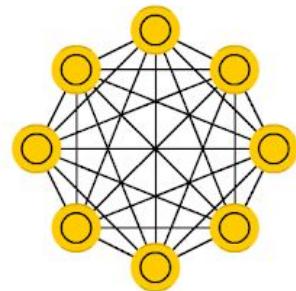
Sparse AE (SAE)



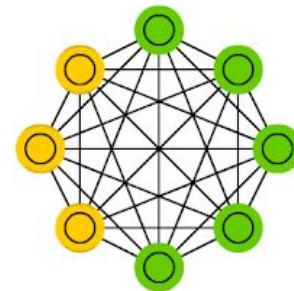
Markov Chain (MC)



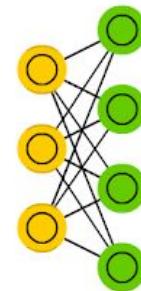
Hopfield Network (HN)



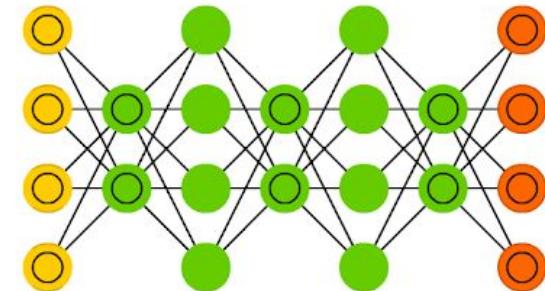
Boltzmann Machine (BM)



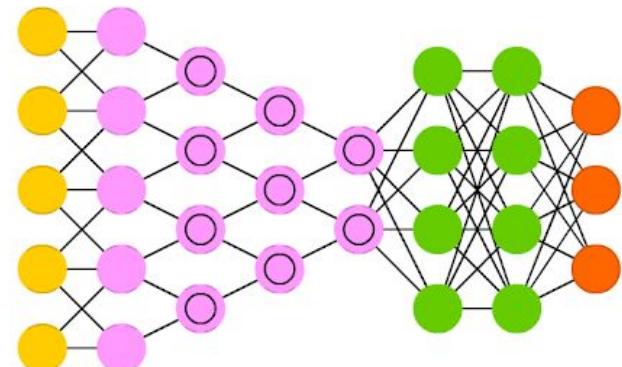
Restricted BM (RBM)



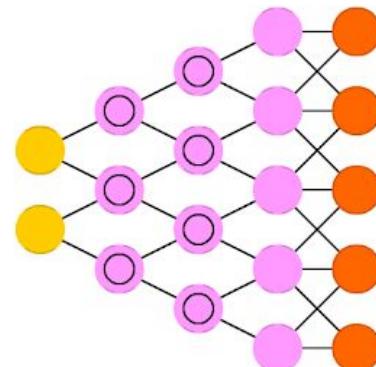
Deep Belief Network (DBN)



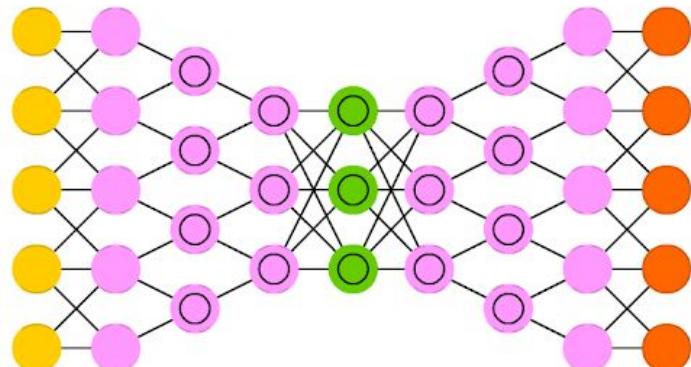
Deep Convolutional Network (DCN)



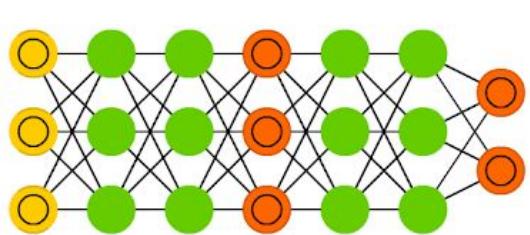
Deconvolutional Network (DN)



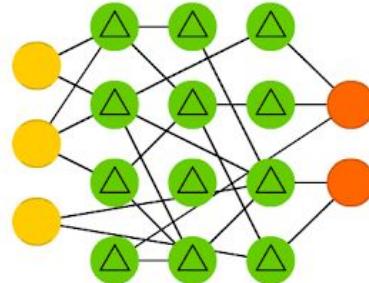
Deep Convolutional Inverse Graphics Network (DCIGN)



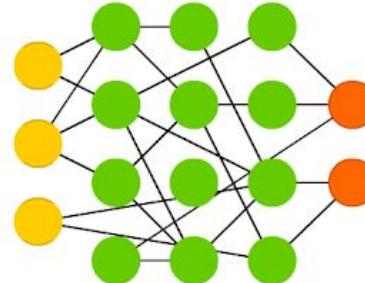
Generative Adversarial Network (GAN)



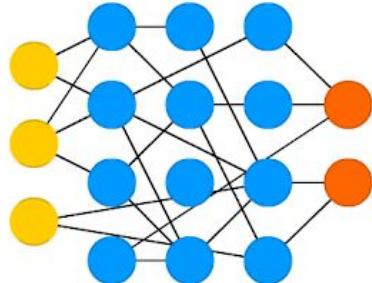
Liquid State Machine (LSM)



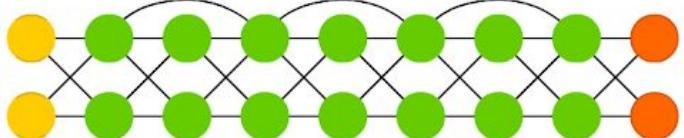
Extreme Learning Machine (ELM)



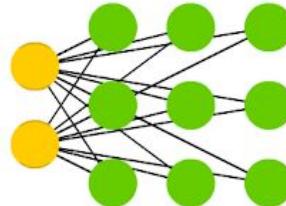
Echo State Network (ESN)



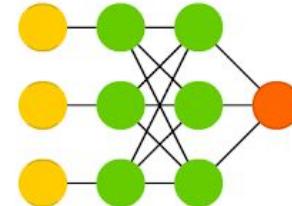
Deep Residual Network (DRN)



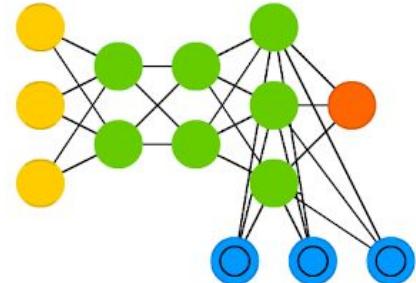
Kohonen Network (KN)



Support Vector Machine (SVM)

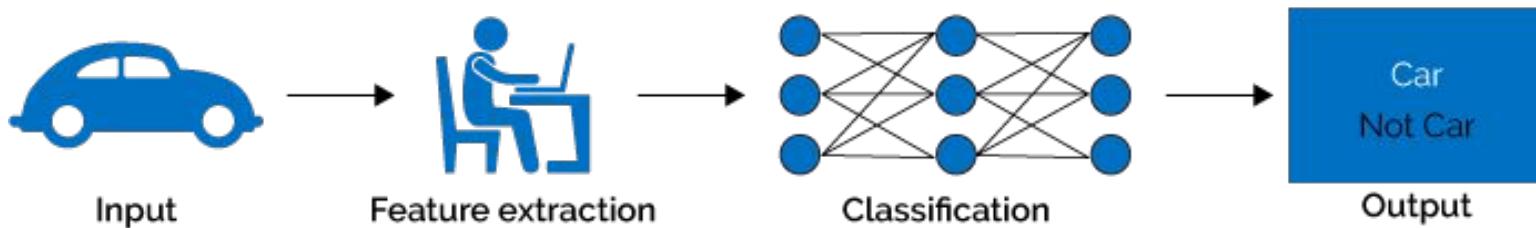


Neural Turing Machine (NTM)

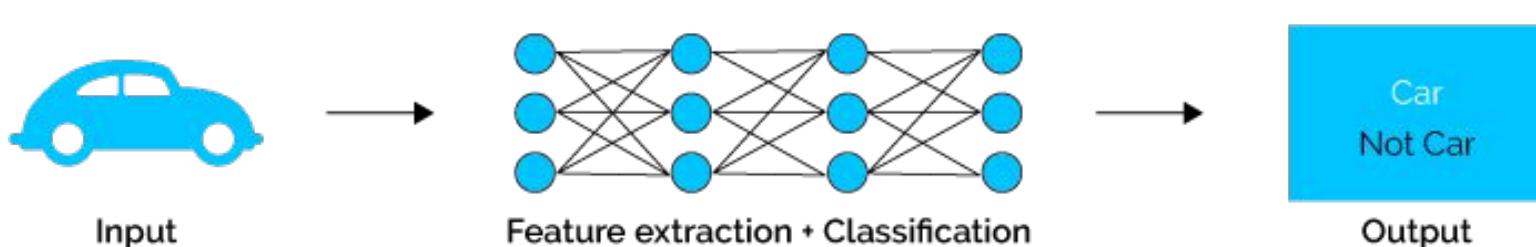


# Extração de Features

## Machine Learning



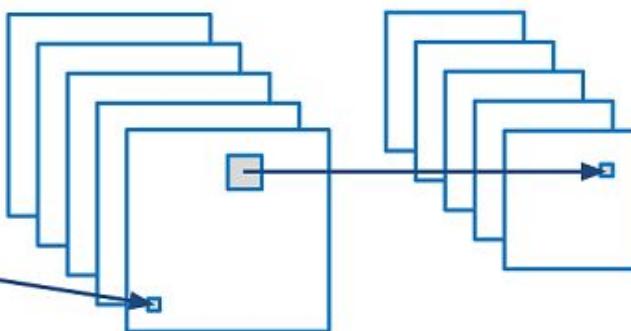
## Deep Learning



# Redes Neurais Convolucionais

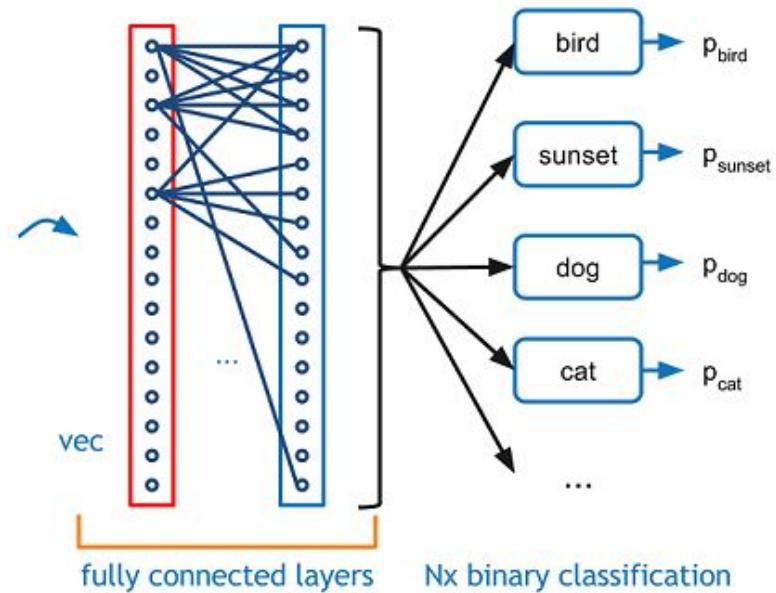


convolution +  
nonlinearity



convolution +  
pooling layers

max pooling



fully connected layers

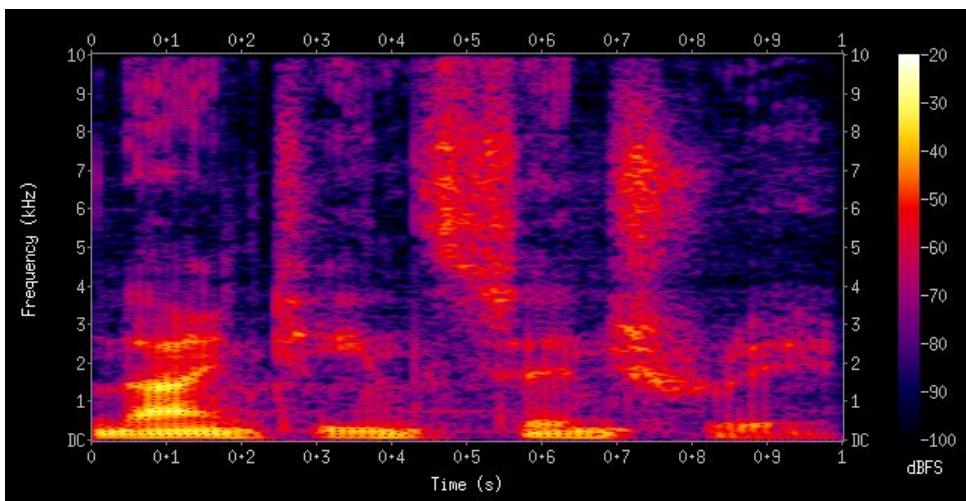
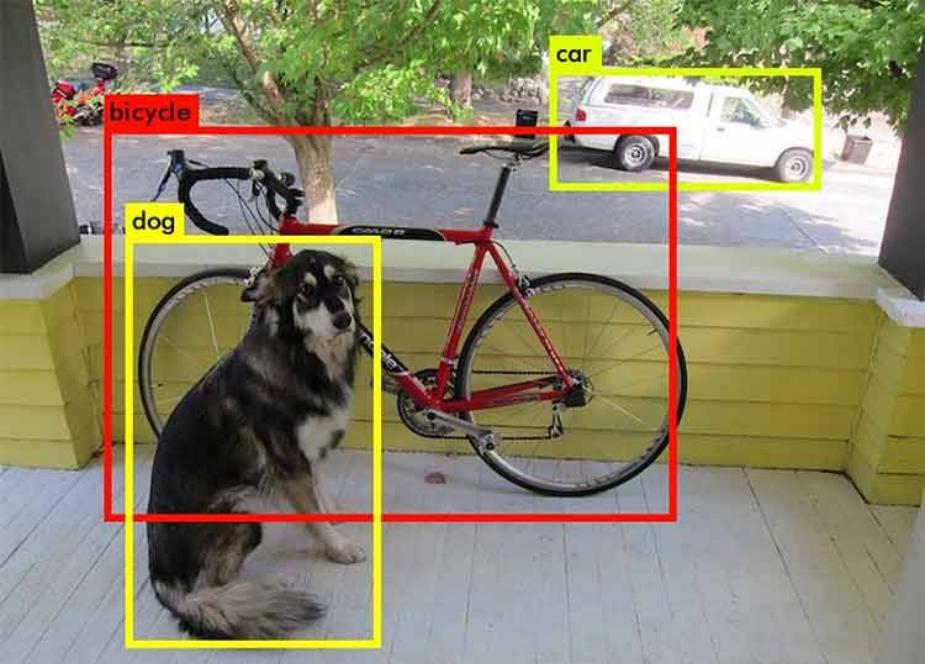
Nx binary classification

# Redes Neurais Convolucionais

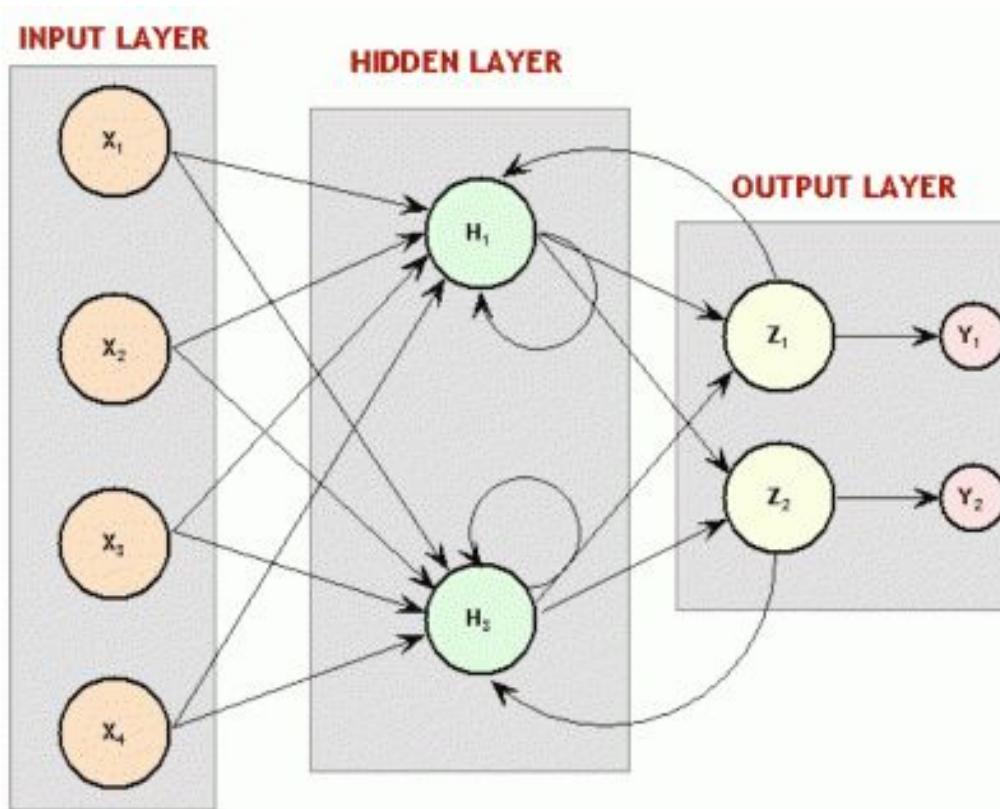
---



# Redes Neurais Convolucionais

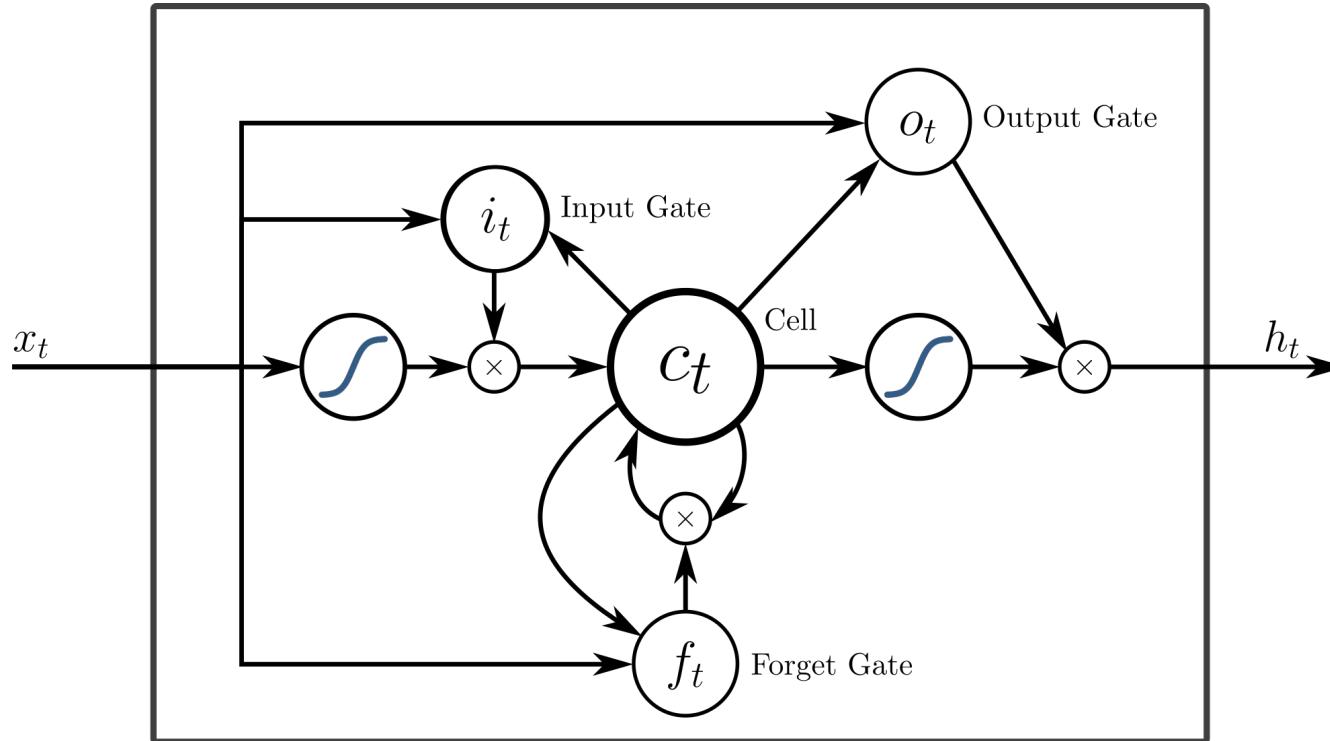


# Redes Neurais Recorrentes

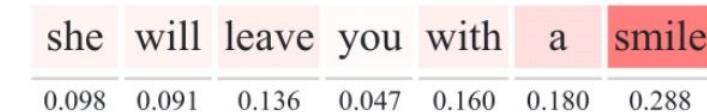
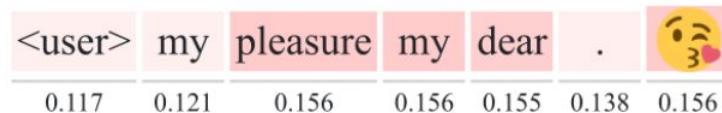
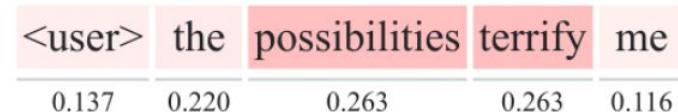
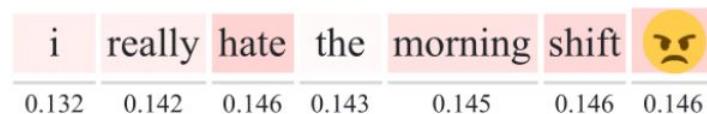
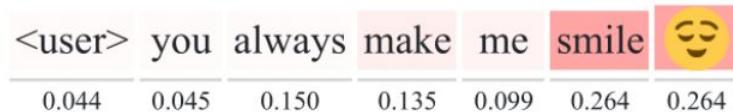
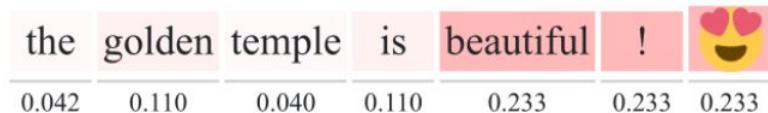
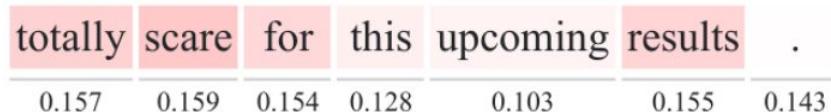
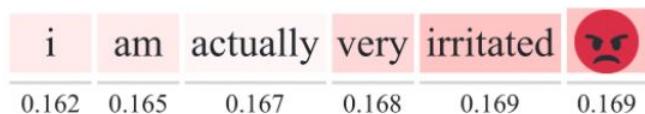


# Long-Short Term Memory (LSTM)

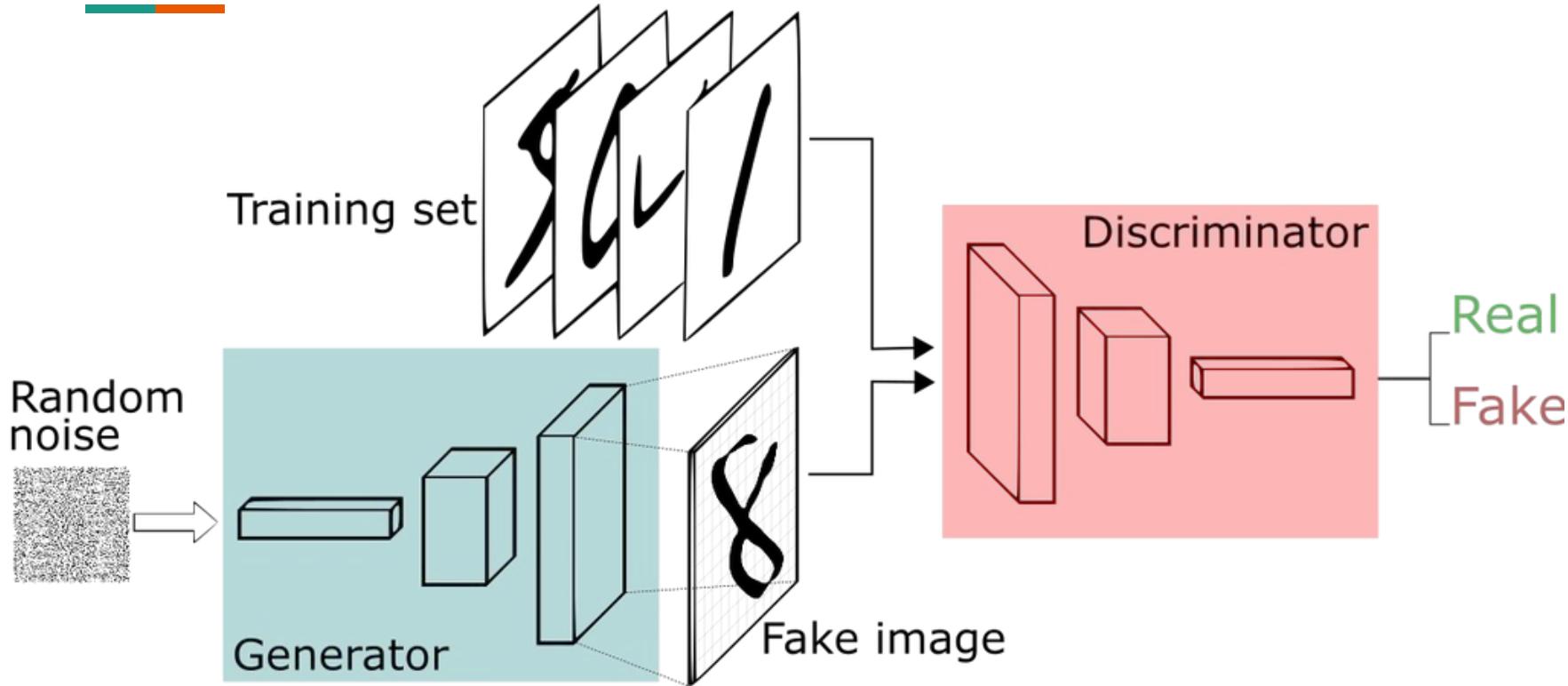
---



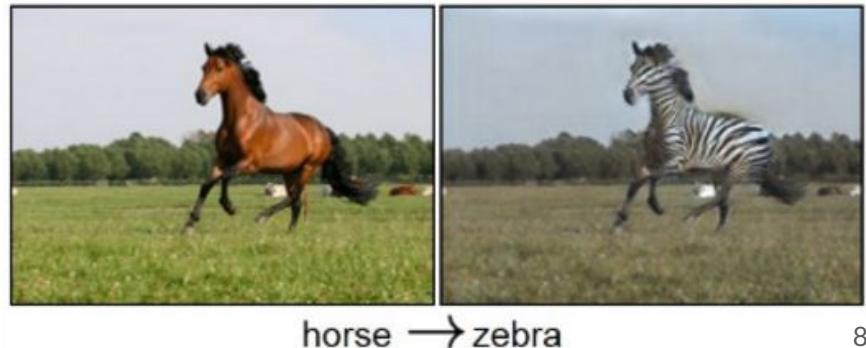
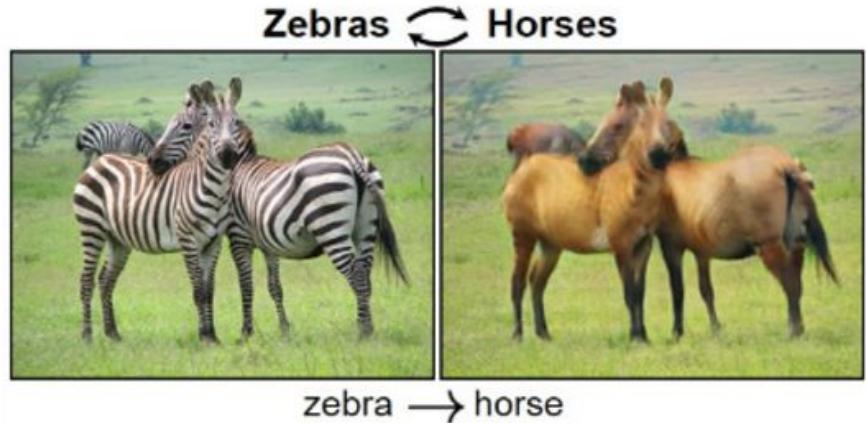
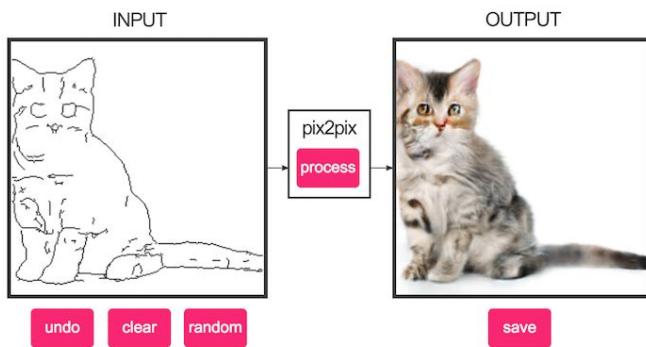
# Redes Neurais Recorrentes e LSTMs



# Generative Adversarial Network



# Generative Adversarial Network



# Testando Redes Neurais **Código**

---

<http://goo.gl/vVULci>

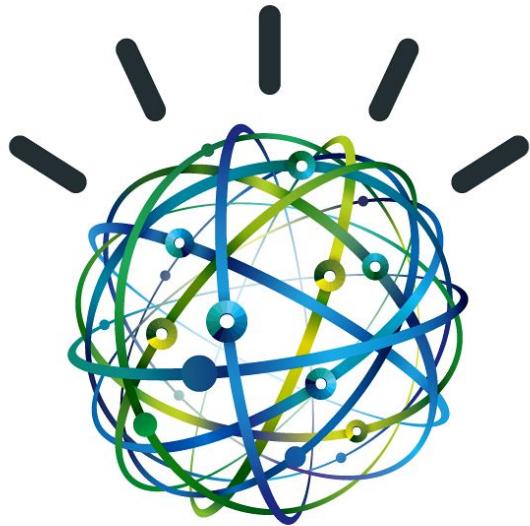
# Limitações das NNs

---

- Grande quantidade de dados
- Poder de processamento (GPUs, TPUs)
- Entendimento da matemática (álgebra linear)

# Ferramentas de alto nível

---



## IBM Watson

# Ferramentas de alto nível

---



Google Cloud AI

Use your own data to train models



TensorFlow



Cloud Machine Learning Engine

Ready to use Machine Learning models



Cloud Vision API



Cloud Speech API



Cloud Jobs API



Cloud Translation API



Cloud Natural Language API



Cloud Video Intelligence API



Coming soon

# Ferramentas de alto nível

---



**Firebase ML Kit  
(Mobile)**

Recurso	No dispositivo	Nuvem
Reconhecimento de texto (OCR, na sigla em inglês)	✓	✓
Detecção facial	✓	✓
Leitura de código de barras	✓	
Marcação de imagens		✓
Reconhecimento de logotipos		✓
Reconhecimento de pontos de referência		✓
Detecção de conteúdo explícito		✓
Pesquisa de imagem semelhante		✓
Inferência de modelo personalizado	✓	

# Ferramentas de alto nível



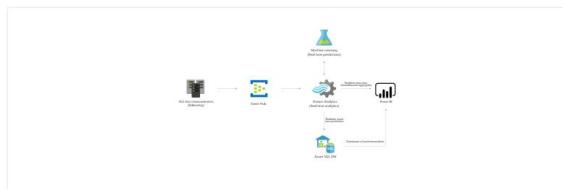
# Azure AI



Image classification with convolutional neural networks

Explore transfer learning, convolutional neural networks, and gradient-boosting decision tree algorithms.

[Learn more >](#)



## Defect prevention with predictive maintenance

Learn how to use Azure Machine Learning to predict failures before they happen with real-time assembly line data.

[Learn more >](#)



Information discovery with deep learning and natural language processing

See how deep learning and natural language processing can be used effectively with the Microsoft AI platform.

[Learn more >](#)



Enterprise Productivity Chatbot

Azure Bot Service can be easily combined with Language Understanding to build powerful enterprise productivity bots, allowing organizations to streamline common work activities by integrating external systems, such as Office 365 calendar, customer cases stored in Dynamics CRM and much more.

[Learn more >](#)

# Revisando Conceitos

# Video Aulas

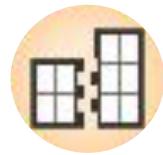
---



simplilearn

# Pré-Processamento no Orange

---



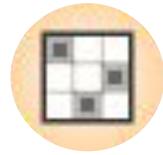
Merge Data



Transpose



Feature Constructor



Impute



Discretize



Color Data

# Datalku

---



## Cloud Machine Learning

<https://hosted-trial.dss-demo.dataiku.com/>

<https://academy.dataiku.com/latest/>

# Microsoft Azure

---



## Azure Machine Learning Studio

<https://studio.azureml.net>

# Fixando Métricas de Classificação

---

Datasets: Qualquer um.

Escolha um dataset do Orange e avalia alguns algoritmos de classificação para ver qual funciona melhor com o dataset escolhido.

- Qual o melhor algoritmo para o seu dataset?

# Material de Classificação

---

- Slides sobre Decision Tree
- Slides sobre Naive Bayes
- Machine Learning 101

# Fixando Classificação

---

Datasets: *Attrition Train* e *Bank Marketing*

Avalia alguns algoritmos de classificação para ver qual funciona melhor com o dataset escolhido. Faça alteração nos parâmetros para ver a melhora das métricas.

- Qual as features mais importantes para o dataset?
- Qual a métrica que deve ser utilizada?

# Material de Regressão

---

- Slides sobre Linear Regression
- Towards Data Science: [Intro](#), [Mais](#)
- [Decision Tree Regressor](#)

# Fixando Regressão

---

Datasets: *HDI* e *Boston Housing*

Estude o dataset (features e target) e proponha um algoritmo de regressão para o problema. Faça variação dos parâmetros.

- Qual as features mais importantes para o dataset?
- Proponha uma visualização dos dados considerando as features importantes.

# Fixando Pré-Processamentos

---

Datasets: *Boston Housing* e *Bank Marketing*

Faça transformações nos dados e verifique a alteração no desempenho dos algoritmos.

- Qual algoritmo mais se beneficia com o pré-processamento?

# Material de Aprendizado Não-Supervisionado

---

- Regras de Associação
- Slides de K-Means Clustering
- Principal Component Analysis

# Fixando Aprendizado Não-Supervisionado

---

Datasets: *Foodmart 2000, Bank Marketing,*

Utilize os algoritmos de não supervisionados para fazer análises dos datasets.

- Qual o agrupamento mais significativo?
- Qual a regra de associação mais relevante?

# Material de Redes Neurais

---

- Slides de Redes Neurais
- Machine vs Deep Learning
- Tutorial de Neural Network

# Fixando Redes Neurais

---

Datasets: *Boston Housing* e *Bank Marketing*

Troque os hiper-parâmetros das redes neurais para encontrar uma topologia que resolve melhor a tarefa.

- Qual a melhor configuração para cada dataset?

# Material de PLN

---

- Slides de PLN
- Pré-Processamento de Texto
- Classificação de Texto

# Fixando PLN

---

Datasets: *Green Tails*

Altere as configurações do pré-processamento do texto para melhorar a predição das classes.

- Qual o melhor pré-processamento?

# Documentação Orange

---



[Documentação](#)

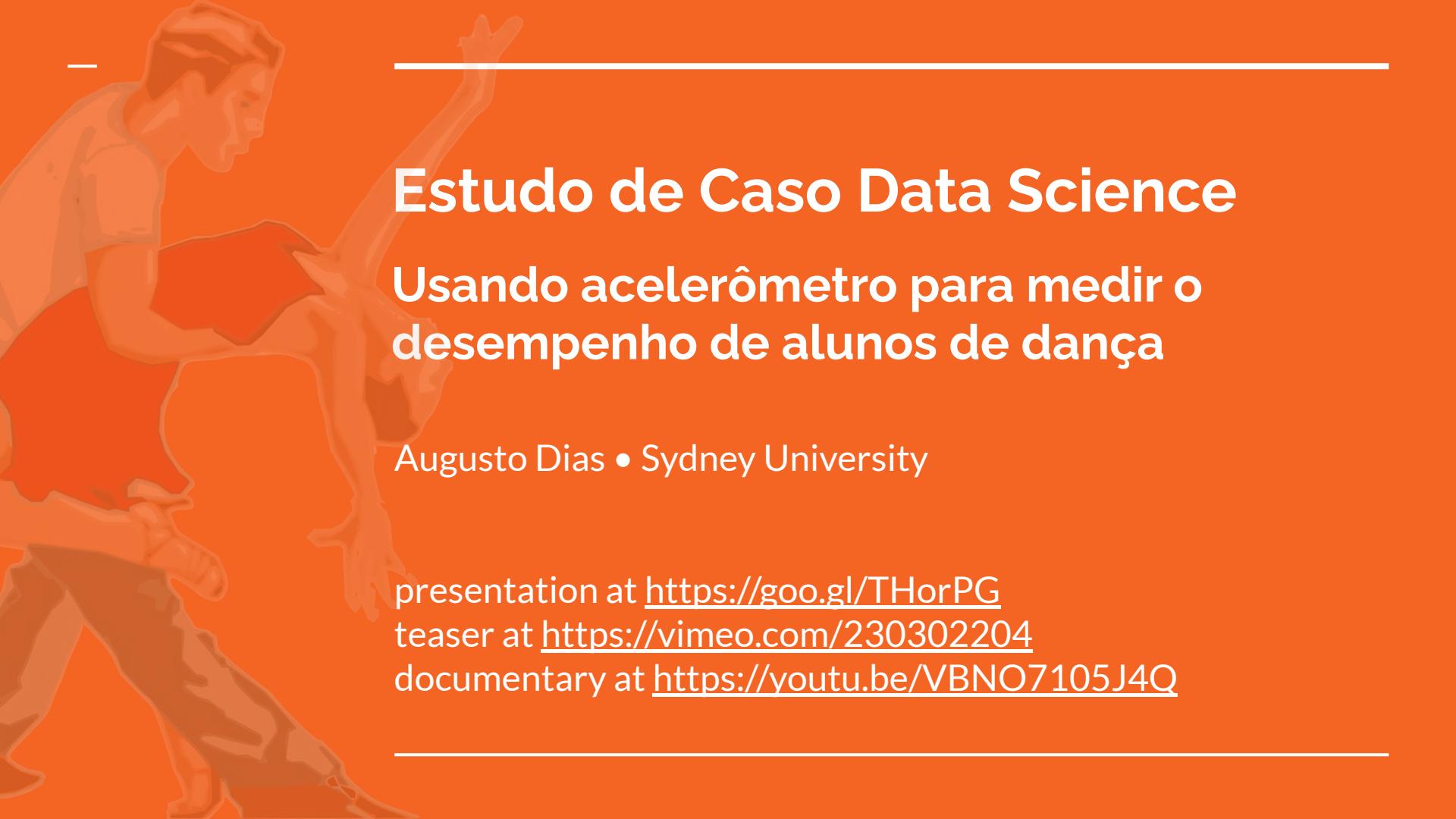
[Canal Oficial no YouTube](#)

# Conteúdo Extra

# Curso de Extensão - PUCRS

---

1º Encontro	2º Encontro	3º Encontro	4º Encontro	5º Encontro	6º Encontro
Conceitos Básicos  Tipos de Aprendizagens	Features vs Target  Classificação  Avaliação de Modelos	Regressão  Agrupamento  Regras de Associação  <b>Resolução de Exercícios</b>	Redes Neurais  Processamento da Linguagem Natural  Pré-Processamento	Ciência de Dados nas Nuvens  Dataiku  Azure ML Studio	Casos de Sucesso BR  <b>Resolução de Exercícios</b>



# Estudo de Caso Data Science

## Usando acelerômetro para medir o desempenho de alunos de dança

Augusto Dias • Sydney University

presentation at <https://goo.gl/THorPG>

teaser at <https://vimeo.com/230302204>

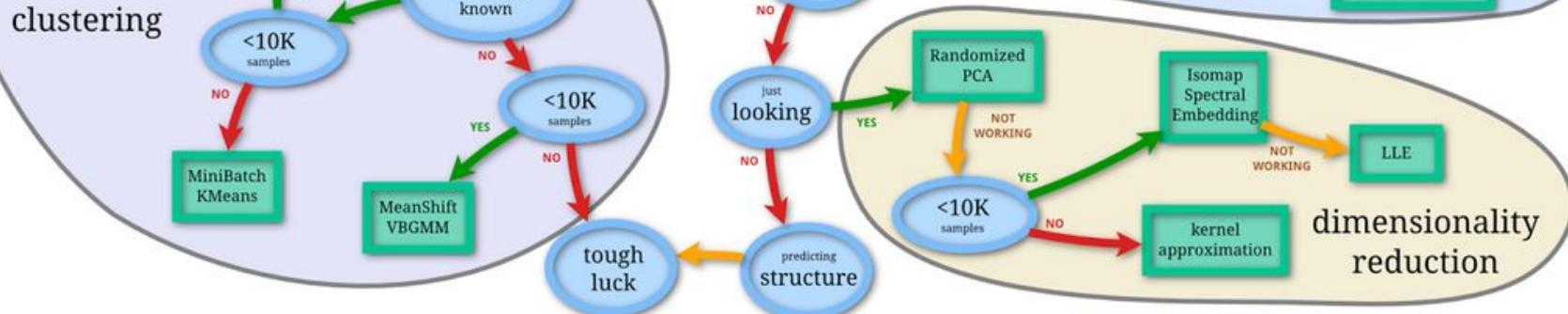
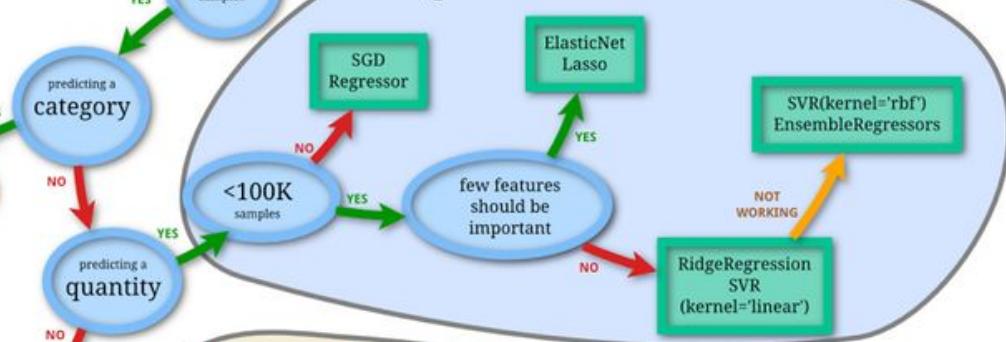
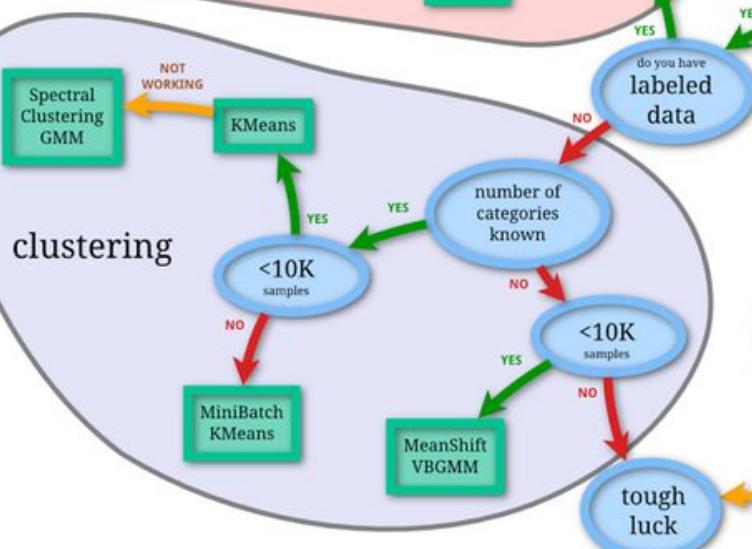
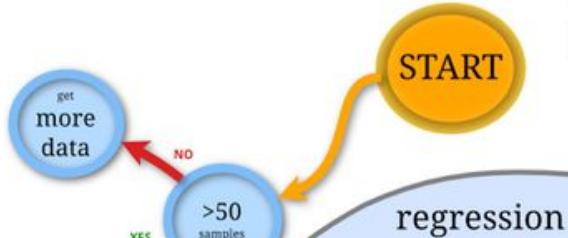
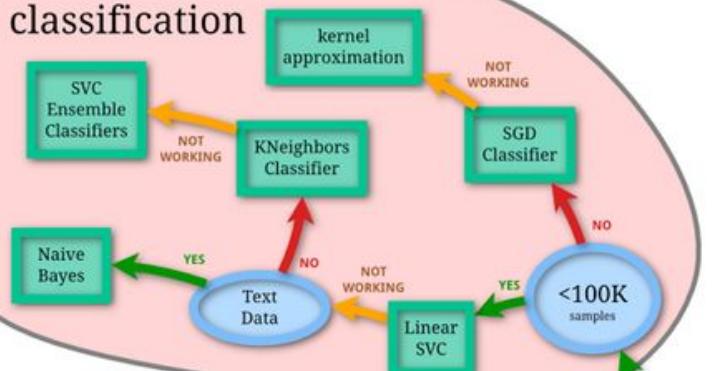
documentary at <https://youtu.be/VBNO7105J4Q>

# Objetivo

---

Visa capacitar os participantes em análise de dados utilizando modelos estatísticos e aprendizagem de máquina.

# scikit-learn algorithm cheat-sheet



# Notebooks da Oficina

---

- Notebook Básico
- Métricas e Regressão Notebook
- Detecção de Outliers

Para todos os links do Colab:

- ao clicar no link, no Google Drive, ir em “Abrir com: Colaboratory”
- dentro do Google Colab, ir em “File”, depois em “Save a copy in Google Drive”
- assim será possível modificar o arquivo e rodar os experimentos

# Outros Notebooks

---

- [Regressão e Busca Exaustiva](#)
- [Processamento da Linguagem Natural](#)
- [Redes Neurais](#)

Para todos os links do Colab:

- ao clicar no link, no Google Drive, ir em “Abrir com: Colaboratory”
- dentro do Google Colab, ir em “File”, depois em “Save a copy in Google Drive”
- assim será possível modificar o arquivo e rodar os experimentos

# Datasets Abertos

---

- <https://www.kaggle.com/>
- <https://www.openml.org/>
- <https://toolbox.google.com/datasetsearch>
- <https://physionet.org/data/> (dados médicos)
- <https://www.cancerimagingarchive.net/> (imagens)

# Outras Palestras (do henrique)

---

- Palestra Nubank ([vídeo](#))
- Ciência de Dados na Saúde ([vídeo](#))
- Data Science para Publicidade
- Aprendizado Não-Supervisionado e o PageRank ([vídeo](#))
- Redes Neurais Self-Attention ([vídeo original](#))

# Cursos Online:

---

- <https://www.kaggle.com/learn/overview>
- <https://www.elementsofai.com/>
- <https://www.datascienceacademy.com.br/pages/cursos-gratis>
- <https://br.udacity.com/course/intro-to-data-science--ud359>
- <https://www.microsoft.com/pt-br/academia>
- <https://online.pucrs.br/pos-graduacao/ciencia-de-dados>
- <https://pt.coursera.org/learn/machine-learning>
- <https://www.pucminas.br/PucVirtual/Pos-Graduacao/Paginas/Ci%C3%A3ncia-de-Dados-e-Big-Data.aspx>

# Vídeos no YouTube

---

- [Big Data - Nerdologia](#)
- [Ciência de Dados - Nerdologia](#)
- [Machine Learning - Nerdologia](#)
- [Linguística Forense - Unabomber](#)
- [Robotização Eleitoral - Estadão](#)
- [Redes Neurais - Nerdologia](#)
- [Código Aberto - Capra](#)
- [Nubank On Stage](#)
- [O futuro do seu emprego - Nerdologia](#)
- [Aprendizado por Reforço - AlphaGo](#)
- [AlphaGo Zero - DeepMind](#)
- [Profissional do Futuro - TED Talks](#)
- [Why Deep Learning Now? - ColdFusion](#)
- [AlphaStar - DeepMind](#)
- [Operação Serenata do Amor](#)

# Documentários

---

- Dilema das Redes Sociais, 2020
- The Great Hack, 2019
- Nova: Está Tudo nos Números, 2018
- AlphaGo, 2017
- The Internet's Own Boy, 2014

# Podcasts no Spotify

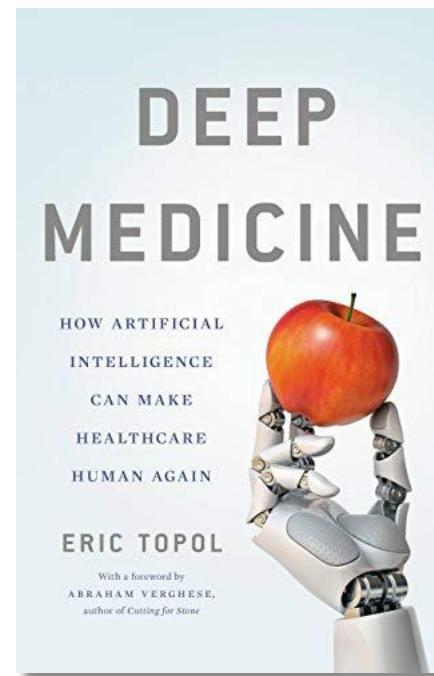
---

- Cabeça de Lab, Magazine Luiza
- Hipsters ponto Tech, Alura

# Livros sobre Ciência de Dados **(para leigos)**

---

- Super Chunchers, Ian Ayres (2007)
- Numeratis, Stephen Baker (2008)
- Deep Medicine, Eric Topol (2019)
- DS para Negócios (2016)
- <https://becominghuman.ai/cheat-sheets-for-ai>



# Casos de Sucesso BR

Visão Única do Cliente	Data Driven Decision para o Ciclo de Vida do Produto	Transformação Omni
<p>Transformar dados em ações</p>  <p><b>INTERAÇÕES COM O CLIENTE</b></p> <ul style="list-style-type: none"><li>• Big data</li><li>• Identificação de não clientes</li><li>• Comunicação personalizada</li><li>• Experiência de compra personalizada</li></ul>	<p>Aplicar tecnologia nas coleções</p>  <p><b>DESENVOLVIMENTO DE PRODUTO</b></p> <ul style="list-style-type: none"><li>• Inteligência artificial<ul style="list-style-type: none"><li>• Algoritmos</li><li>• Predição de vendas</li></ul></li><li>• Agilidade nas decisões</li><li>• Sortimento / Distribuição<ul style="list-style-type: none"><li>• Estoques menores</li></ul></li></ul>	<p>Criar uma experiência de compra única</p>  <p><b>INTEGRAÇÃO DE CANAIS</b></p> <ul style="list-style-type: none"><li>• Uso de estoques das lojas<ul style="list-style-type: none"><li>• Delivery mais rápido</li></ul></li><li>• Equalização dos canais de vendas</li><li>• Conveniência, mobilidade e flexibilidade<ul style="list-style-type: none"><li>• Experiência fluida e encantadora</li></ul></li></ul>

## Status das Iniciativas



- Data lake com dados unificados
- Acesso disponível à áreas de negócio para exploração de dados
- Início da utilização de dados unificados para régua de campanhas de marketing
- Central única de atendimento com disponibilidade de todas informações dos clientes em acesso único
- Motor de recomendação Omni disponível no dispositivo móvel de loja, a partir do perfil do consumidor – 2S20



- Mais de mil itens sendo alocados por IA, sem intervenção humana
- Representatividade de +8% das vendas totais – aumento gradual em 2020
- Receita incremental de 12%, com redução de 18% da necessidade de estoques para esses itens
- Utilização de IA para predição de vendas
- Piloto de captura de tendências de moda a partir de IA



- RFID implementado em 100% das lojas do Brasil
- Testes do Ship from Store
- Retire em loja em 100% das unidades e representando 36% das vendas do e-commerce em dezembro
- Retire em loja utilizando estoque de lojas em 75 unidades – rollout para todo parque em março
- Venda Móvel através do dispositivo específicos em 100% das lojas
- Venda Digital do E-commerce em 100% das lojas
- Caixas de autoatendimento
- Pague Digital - checkout a partir do dispositivo do cliente
- Lockers em lojas
- Same Day Delivery (RJ) e Next Day Delivery (SP)

O que é preciso  
para ter cabeça de lab?

**O NOSSO PODCAST  
JÁ ESTÁ NO AR!**



**luizalabs<sup>o</sup>**

PodCast: <https://open.spotify.com/show/6jYcj4oQ31J85jGhbiRkK>



# Dados **explícitos** e **implícitos**

## Explícito

Dar nota a um produto

Curtir um produto

## Implícito

Frequência de visualização

Adição no carrinho

Clique em foto

Adição nos favoritos

Tempo no produto

# Não-personalizada e Personalizada

## Não-personalizada

Comportamentos Gerais

Diferentes usuários vêem o mesmo conjunto de itens.

Não leva em conta o comportamento individual.

## Personalizada

Comportamentos pessoais

Diferentes interações de cada usuário.

Combinação única entre cliente e produto.

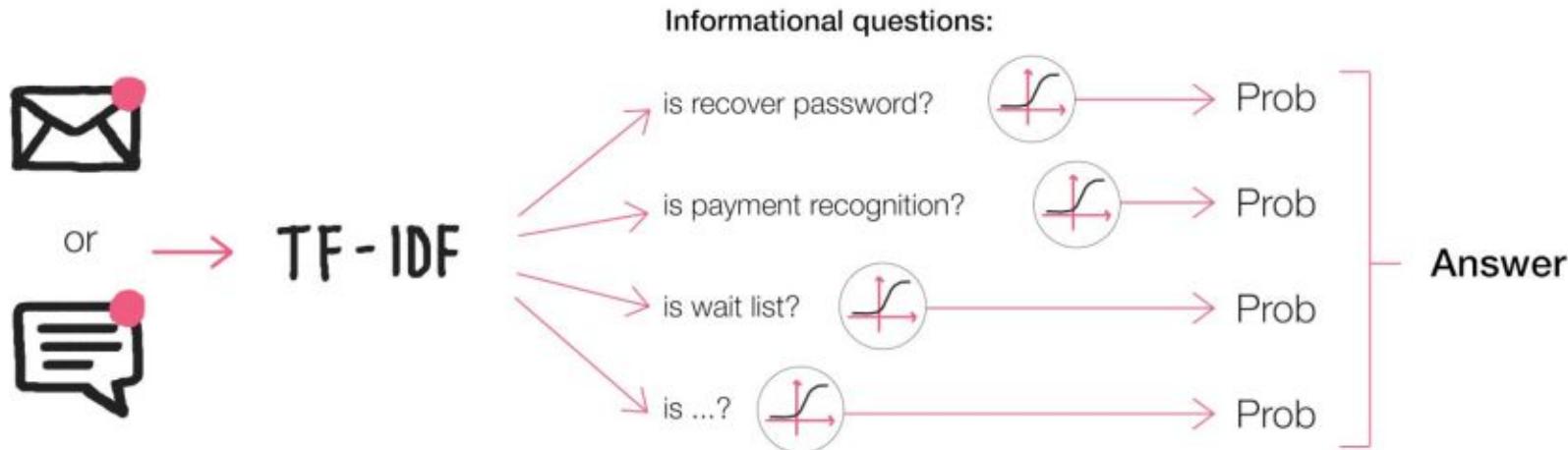
# Data Science em uma instituição financeira moderna

Sandor Tucakov Caetano  
@SandorCaetano  
[sandor.caetano@gmail.com](mailto:sandor.caetano@gmail.com)

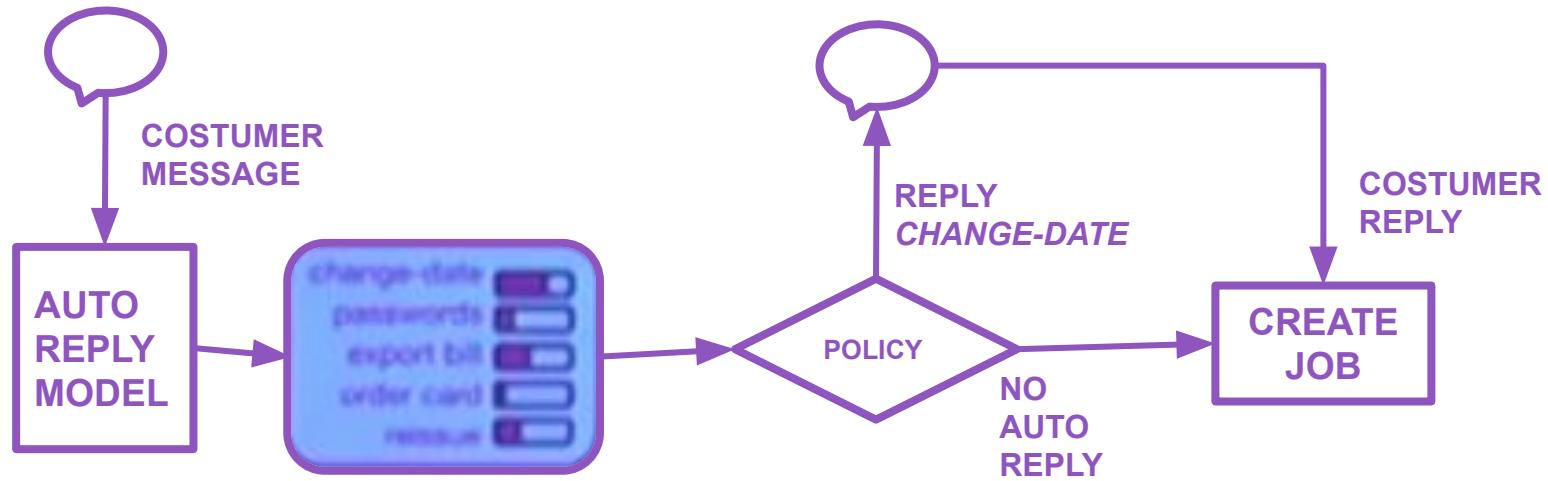
Fonte: <https://pt.slideshare.net/Nubank/data-science-em-uma-instituicao-financeira-moderna>  
Videos: <https://www.youtube.com/channel/UC5yS6v2umolXx8TSJsUEBKg/videos>



# Auto-reply models



# Chat Flow



# Technology platform | Data analytics and AI

Understand customers and leverage the business.



## Generated value

Value capture with AI and **analytics** has been increasing

(base 100)

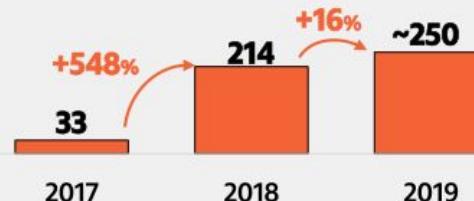


Average ROI **1500%**

## Training and capacity building

Human capital training and knowledge dissemination

Number of data scientists at Itaú Unibanco:



Itaú Unibanco Analytics Training Program:

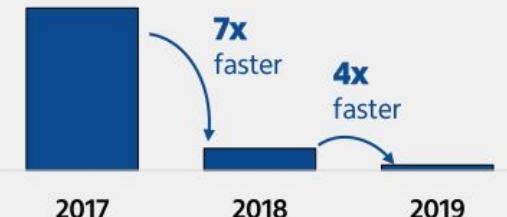
- 7 months long
- 400h training

Partnerships:



## Analytics Ops

Efficient strategy for model deployment



**VIRTUAL ASSISTANT**  
Artificial intelligence  
algorithm is retrained  
every week

# Usamos a tecnologia para resolver problemas reais

Raciocínio tradicional: **visão de oferta**



Raciocínio moderno: **visão de demanda**



**tempo**  
🕒🕒🕒

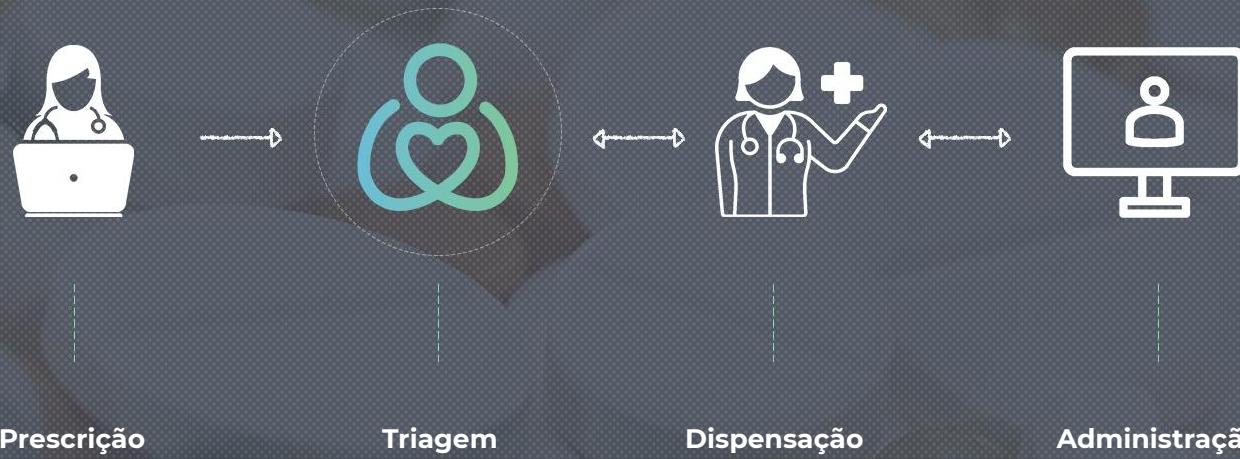
**desembolso**  
\$\$\$

**tempo**  
🕒🕒🕒

**desembolso**  
\$\$\$\$



# Fluxo de Medicamentos



<http://www.lauranetworks.com>





OI, SOU A LAURA!

ALERTAS

