

Data Science Academy

Machine Learning



Seja bem-vindo!





Data Science Academy

www.datascienceacademy.com.br



Data Science Academy

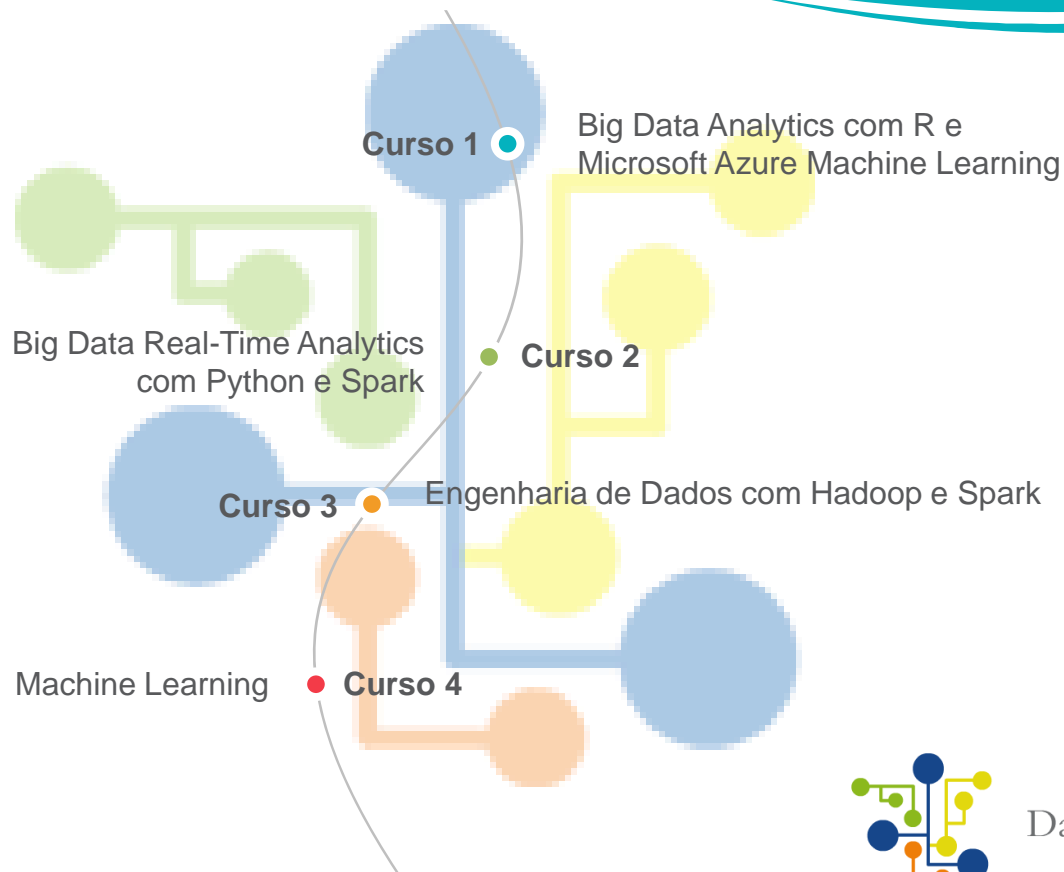


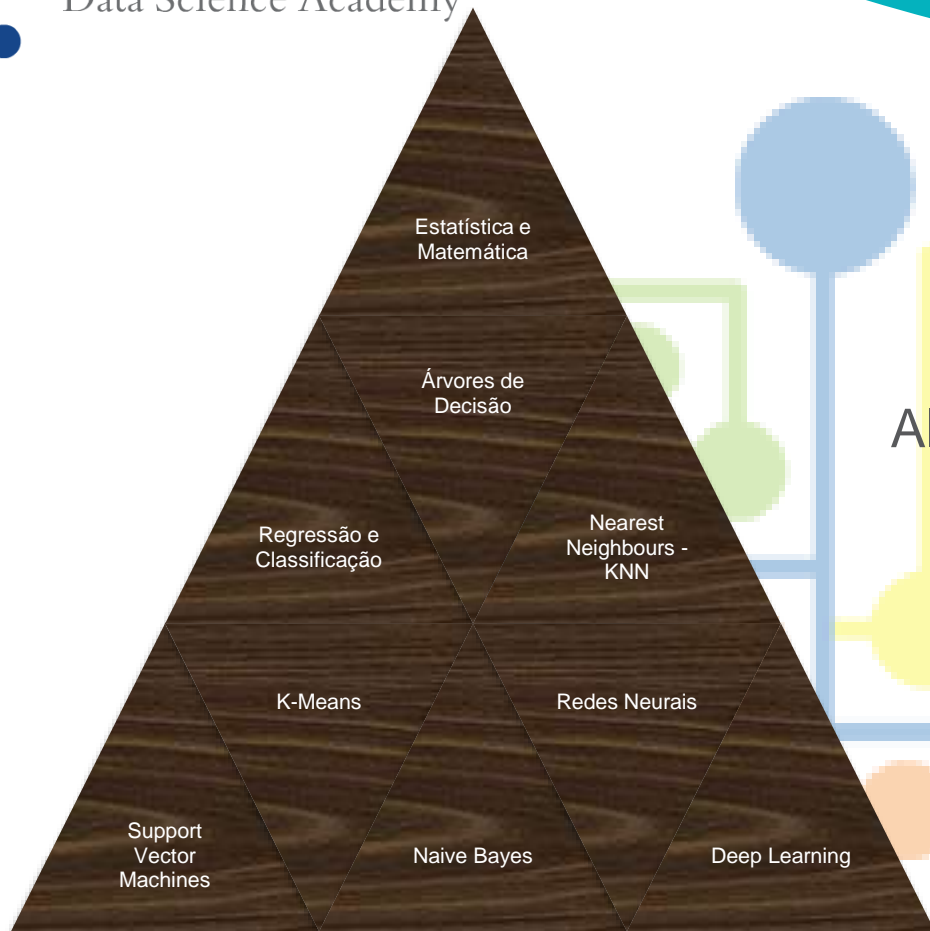
“

Machine Learning

”

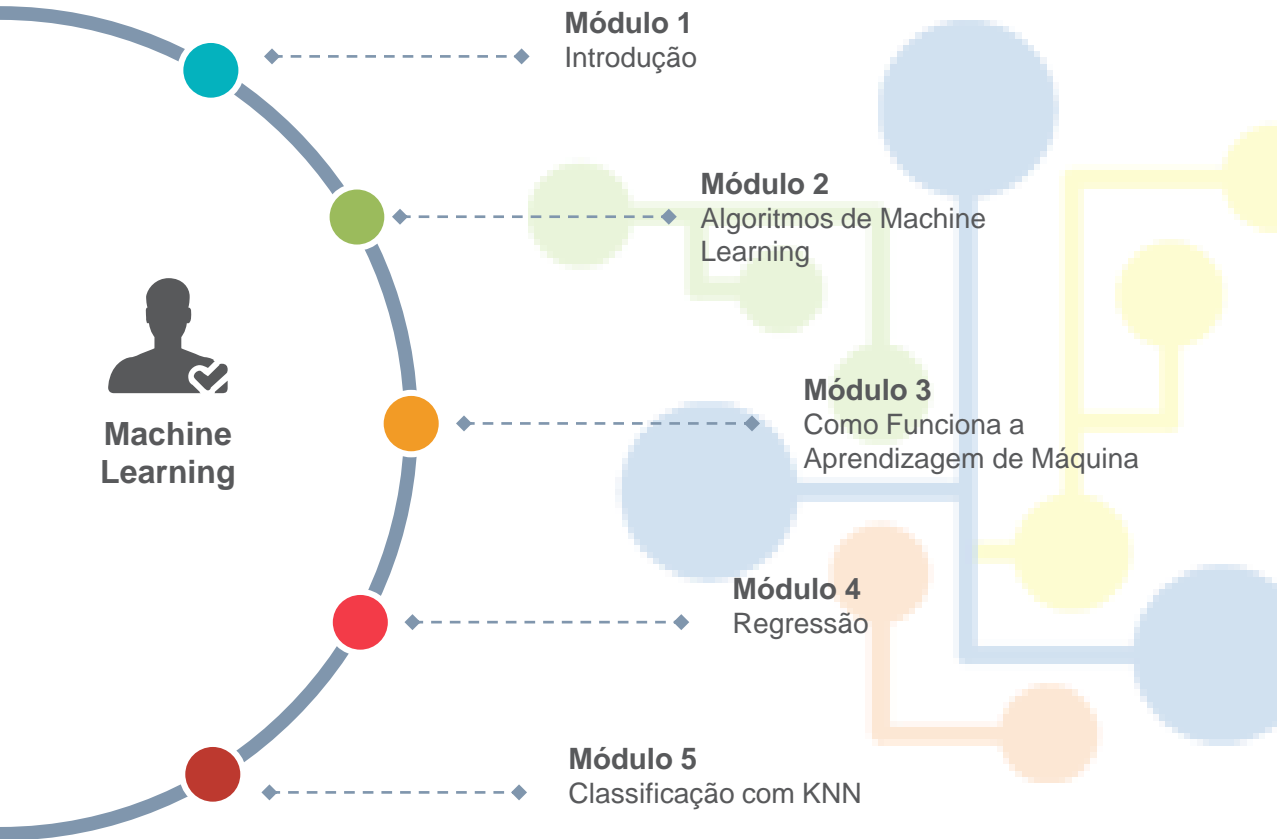


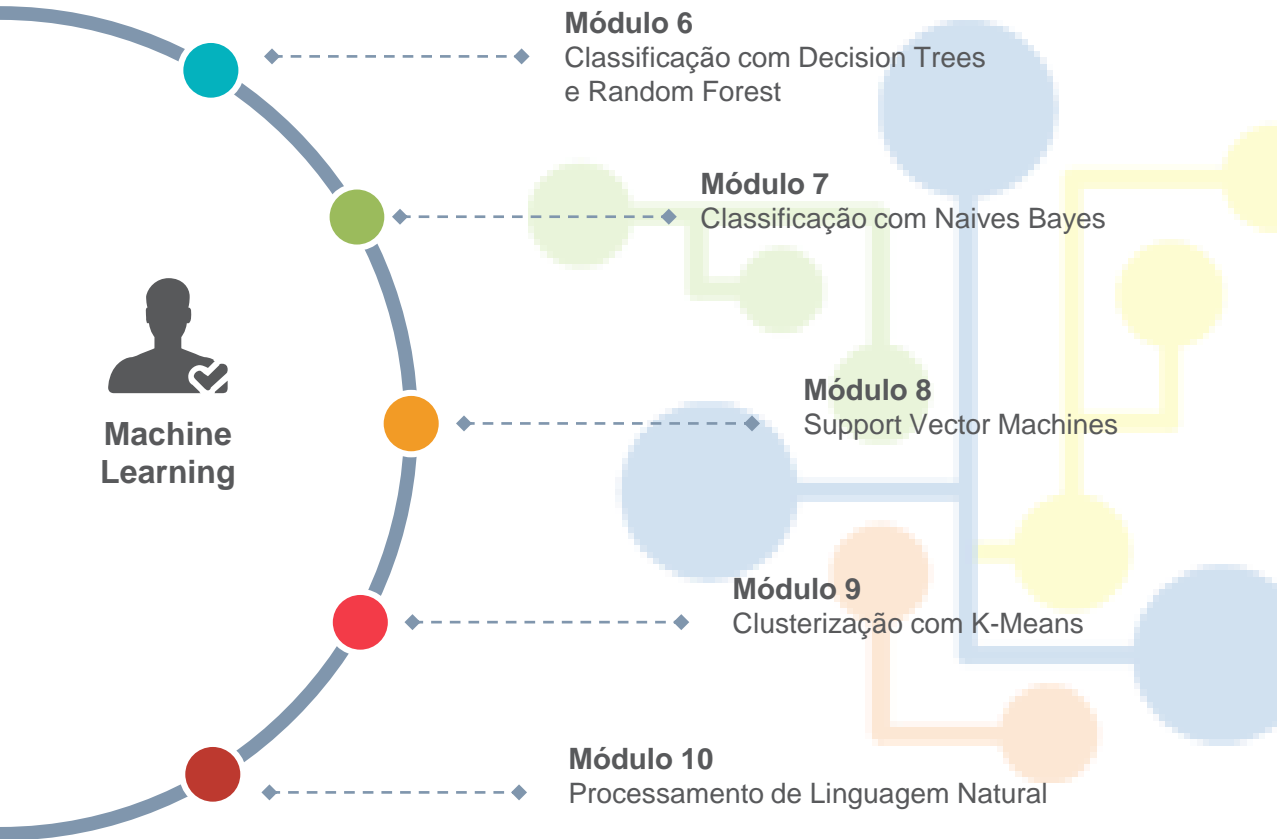


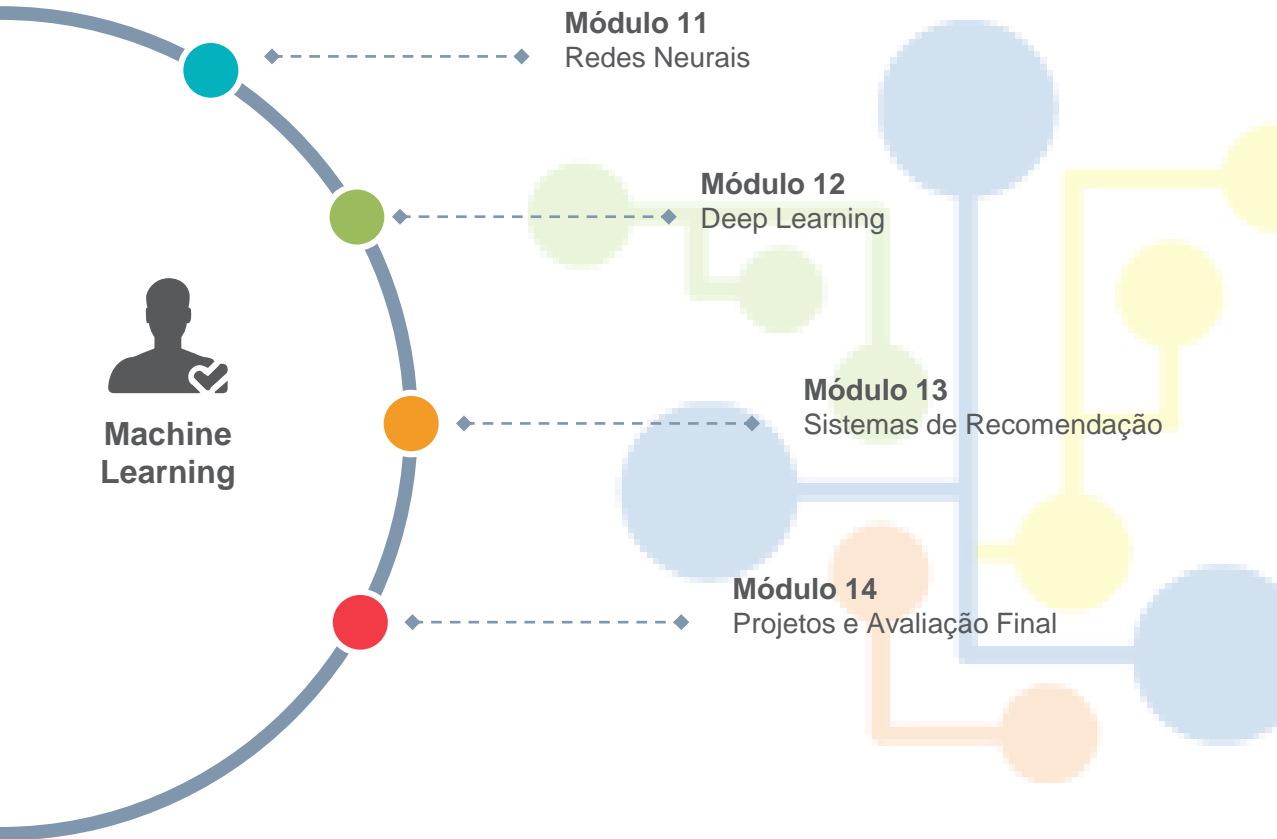


Algoritmos de Machine Learning



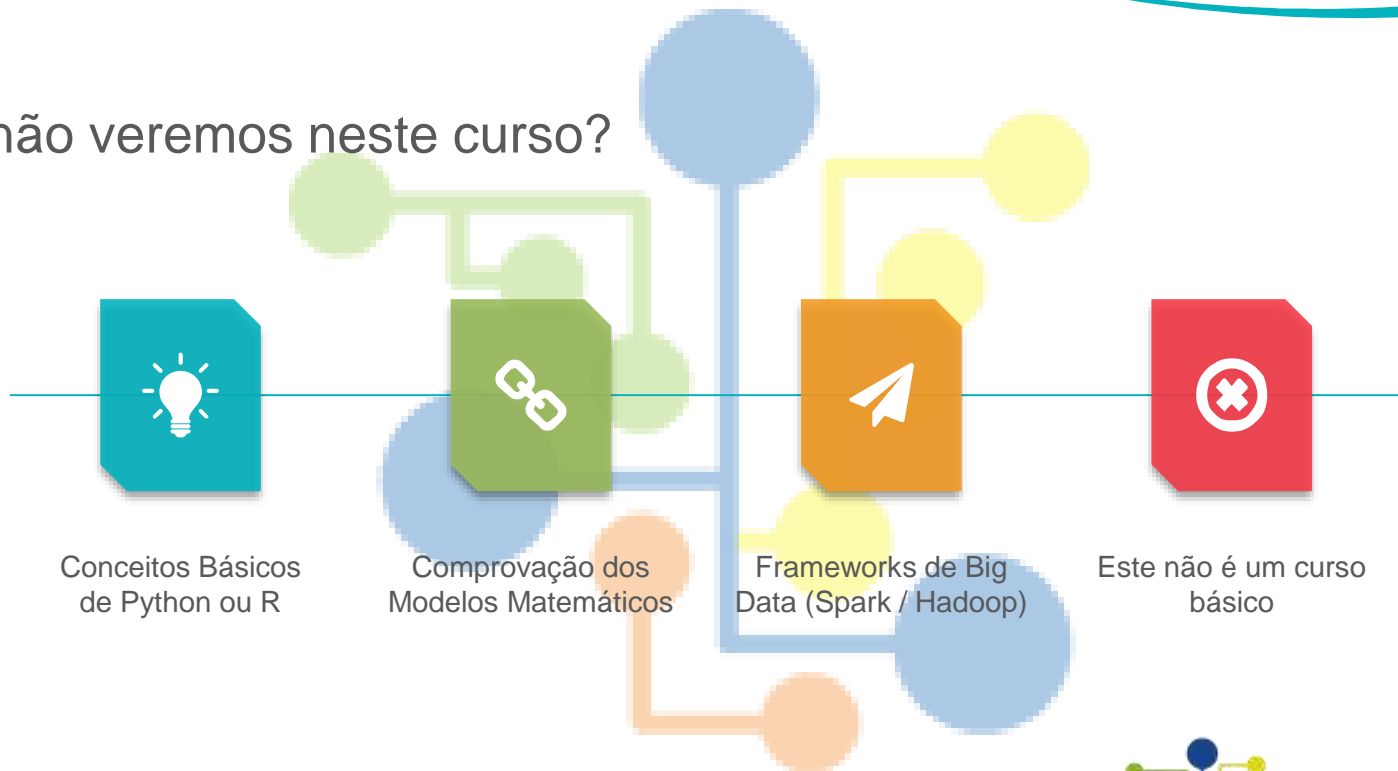


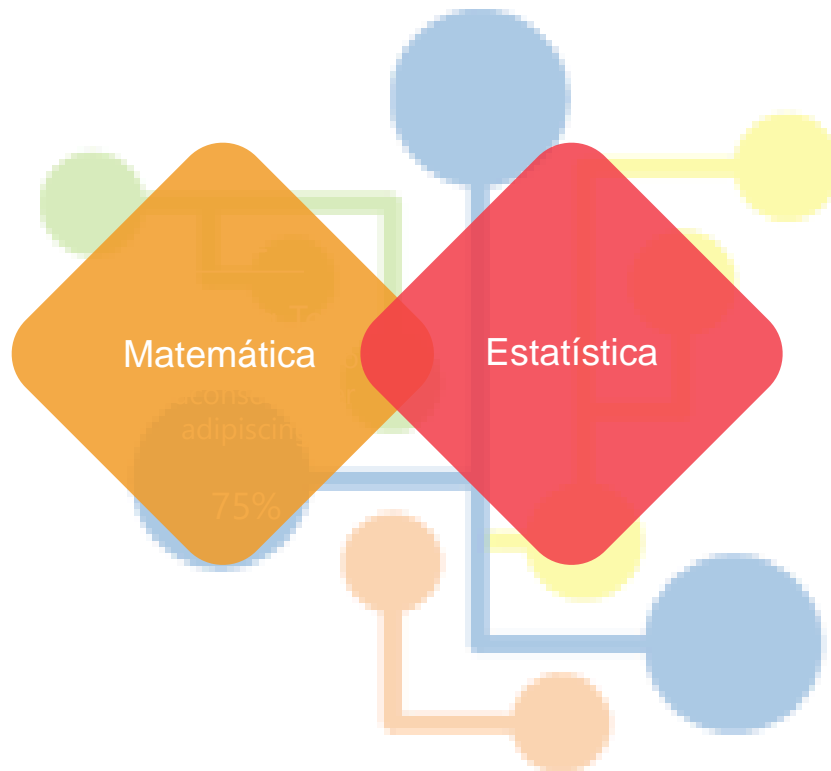






O que não veremos neste curso?







Pré-requisitos

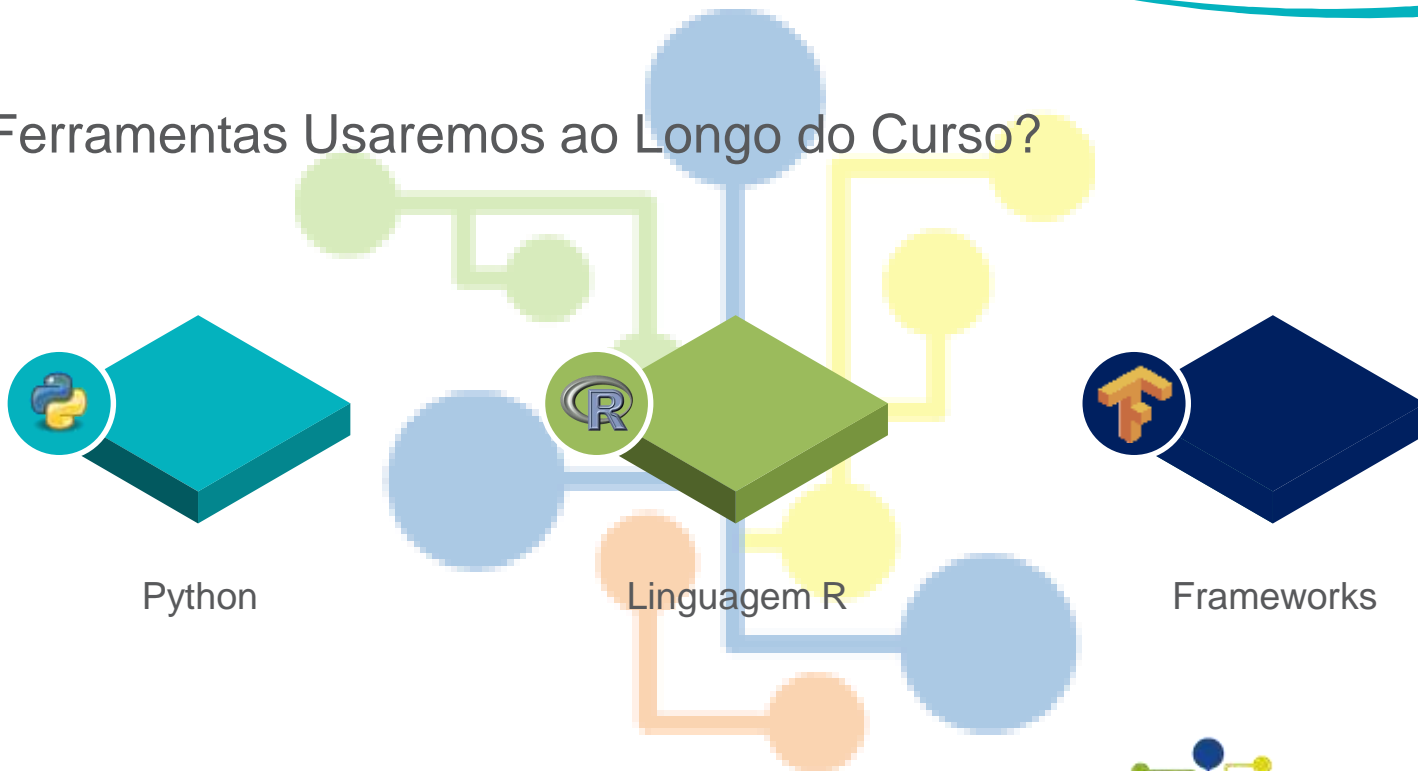


- Python Fundamentos para Análise de Dados
- Big Data Real-Time Analytics com Python e Spark
- R Fundamentos para Análise de Dados
- Big Data Analytics com R e Azure Machine Learning
- Big Data Fundamentos
- Introdução à Ciência de Dados





Quais Ferramentas Usaremos ao Longo do Curso?



Python

Linguagem R

Frameworks





Dedicação

6 a 8 horas por semana



Comunicação

Utilize nossos canais de comunicação



Prática

Você terá acesso a todos os scripts comentados linha a linha

Recomendações

Lembre-se:

Seu aprendizado também depende de você!





O SEGREDO
do seu sucesso
esta na constância
do seu ESFORÇO





Objetivos ao fim deste curso



100%

Desenvolver o processo de modelagem de dados para Machine Learning





Objetivos ao fim deste curso



100%

Conhecer o principais algoritmos de Machine Learning, suas aplicações e diferenças





Objetivos ao fim deste curso



100%

Aprender técnicas de
Machine Learning e
Processamento de Dados





Objetivos ao fim deste curso



100%

Aplicar as técnicas de
aprendizado de máquina
e desenvolver modelos
preditivos





Método de Ensino

Exposição Teórica
Exposição Prática
Exercícios e Quizzes



E-books e Manuais
Bibliografia, Referências e Links Úteis
Scripts





Projetos



Projeto 1 – Implementando um Classificador de Spam com Naïve Bayes



Projeto 2 – Construindo um Sistema de Recomendação de Filmes



Projeto 3 – Criando um Modelo de Machine Learning para Retorno Sobre Investimentos



Projeto 4 – Aplicando Machine Learning para Otimizar o Sistema de Voos de um Companhia Aérea



Projeto 5 – Análise SVM para Prever a Força do Real em Relação a Outras Moedas



Projetos completos, com especificação, solução e documentação, além dos scripts usados para criação dos modelos preditivos

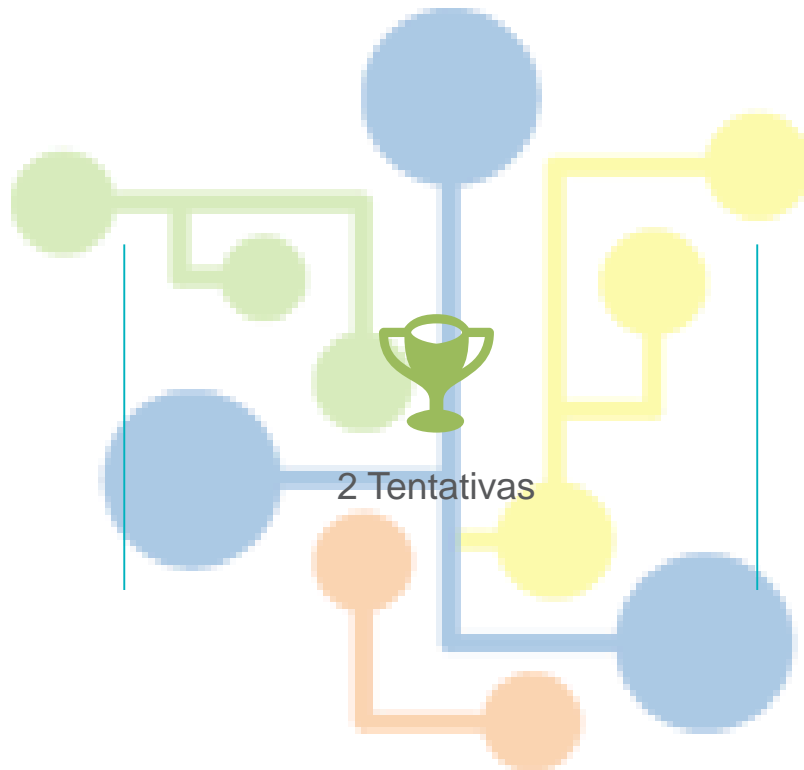




Avaliação Final



50 Questões



2 Tentativas



70% de Aproveitamento





Acesse o Curso do Smartphone ou Tablet com
nossas Apps para iOS e Android



A group of people are shown from the chest up, clapping and smiling. On the left, a man in a light blue button-down shirt is clapping. In the center, a woman with dark hair tied back, wearing a grey t-shirt, is smiling broadly and clapping. To her right, another woman is partially visible, also clapping. On the far right, a man in a white shirt is clapping. The background is a blurred indoor setting. A large teal diagonal overlay covers the left half of the image, with the text 'Divirta-se' in white.

Divirta-se



Data Science Academy

O que é Aprendizado de Máquina?



Data Science Academy

A banner for the 'Formação Cientista de Dados' (Data Scientist Training) course. It features a central image of a person in a white shirt and tie pointing at a laptop screen displaying data visualizations. The background is dark with a subtle pattern of hands. Below the main image are four colored boxes: teal for 'Vídeos Interativos', blue for 'E-books', purple for 'Projetos', and olive green for 'Certificado'.

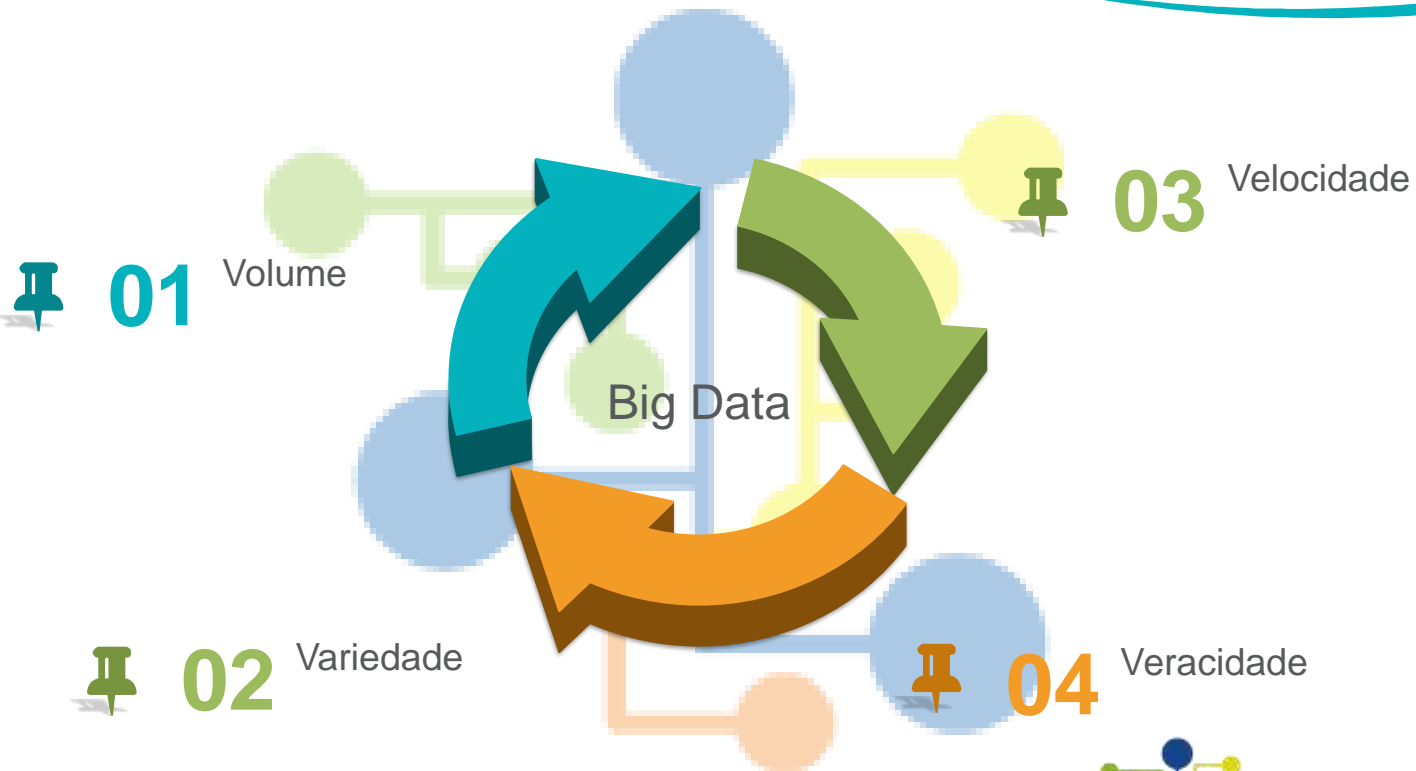
Formação Cientista de Dados

Matricule-se já

Vídeos Interativos 296 horas de curso	E-books Com todo o conteúdo dos cursos	Projetos 26 projetos profissionais	Certificado Certificado de conclusão
---	--	--	--

- Big Data Analytics com R e Azure
- Big Data Real-Time Analytics com Python e Spark
- Engenharia de Dados com Hadoop e Spark
- **Machine Learning**
- Business Analytics
- Visualização de Dados







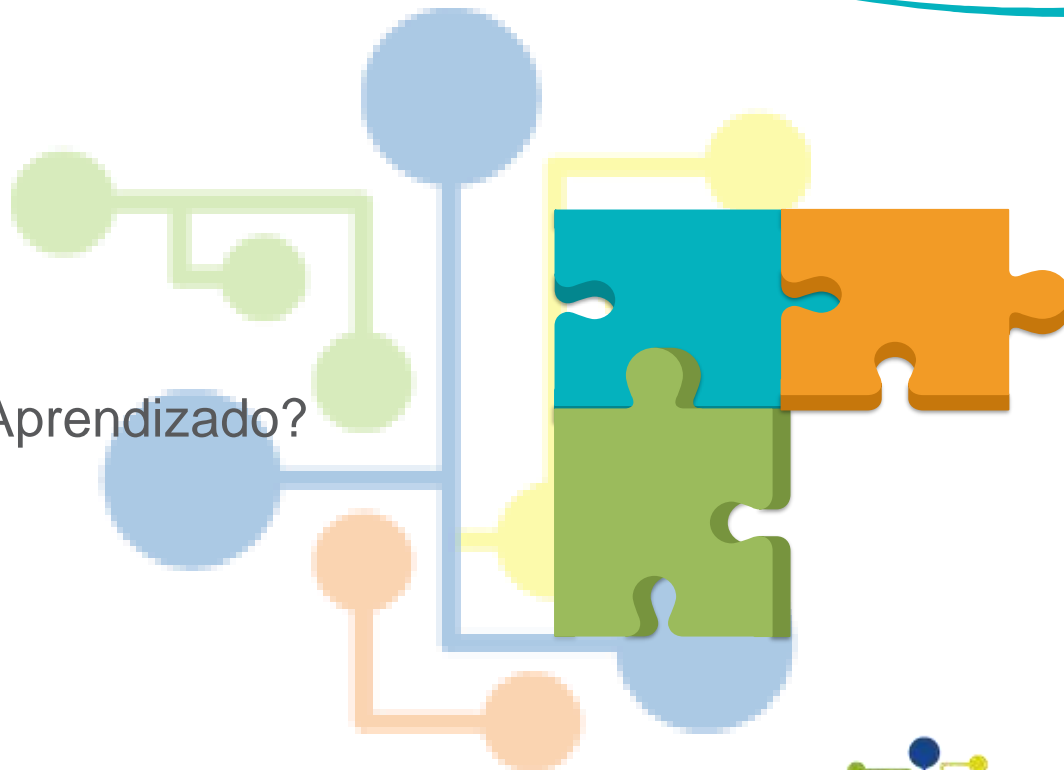
Aprendizagem na Era do Big Data

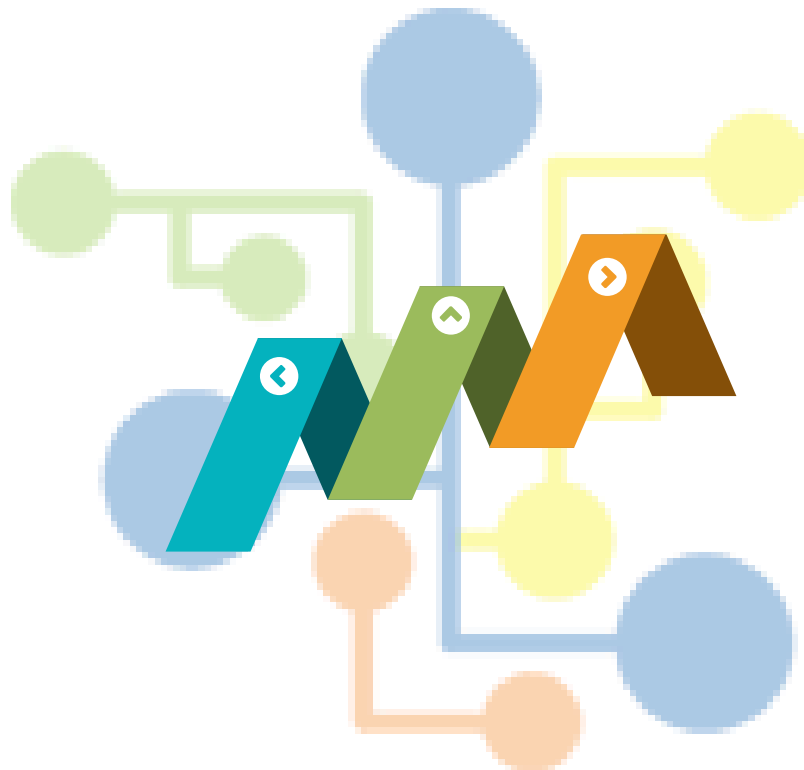
- Grande Volume de Dados
- Evolução das Técnicas Analíticas
- Análise de Dados em Tempo Real
- Desenvolvimento de Aplicações Inteligentes

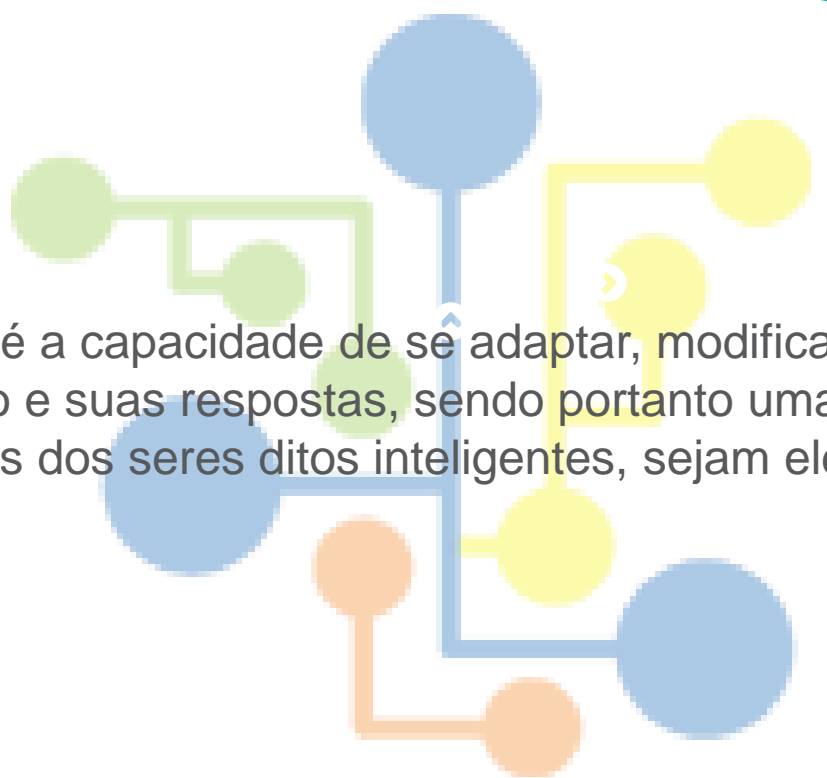




O que é Aprendizado?

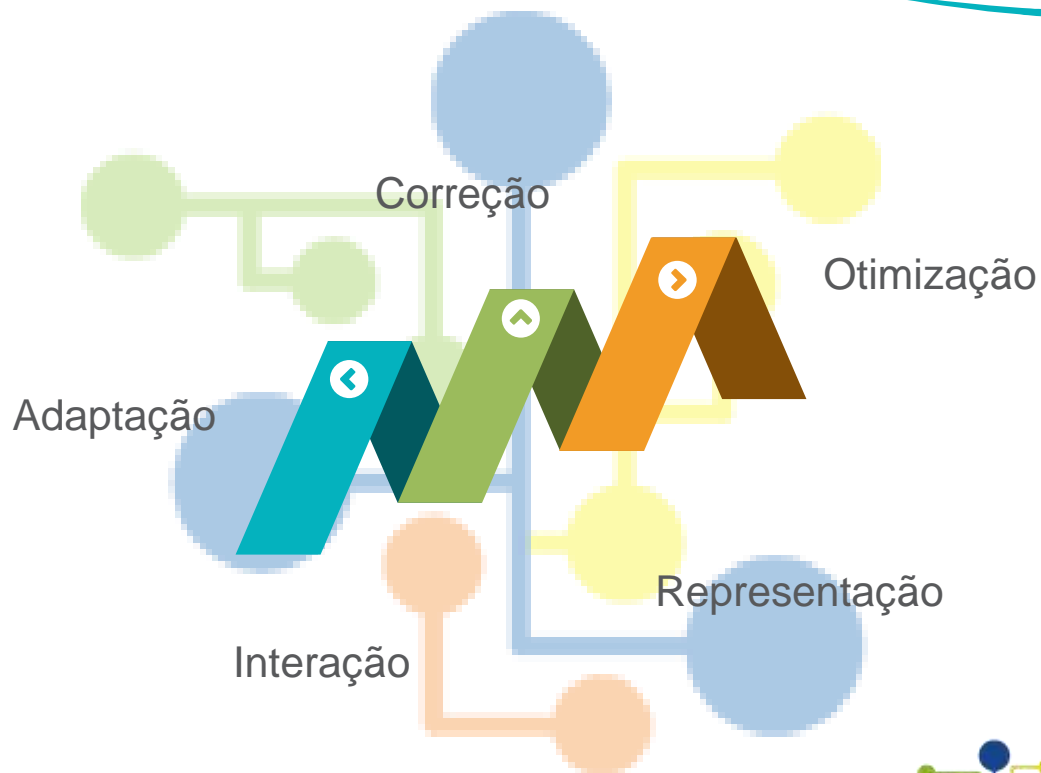




A large, faint, stylized graphic in the background consisting of interconnected nodes and lines in blue, green, yellow, and orange, resembling a neural network or data flow diagram.

Aprendizado é a capacidade de se adaptar, modificar e melhorar seu comportamento e suas respostas, sendo portanto uma das propriedades mais importantes dos seres ditos inteligentes, sejam eles humanos ou não







Percebeu a semelhança do processo de aprendizado de seres humanos e através de algoritmos de Machine Learning?





Já podemos então definir
Aprendizado de Máquina





Machine Learning é um subcampo da Inteligência Artificial que permite dar aos computadores a habilidade de aprender sem que sejam explicitamente programados para isso

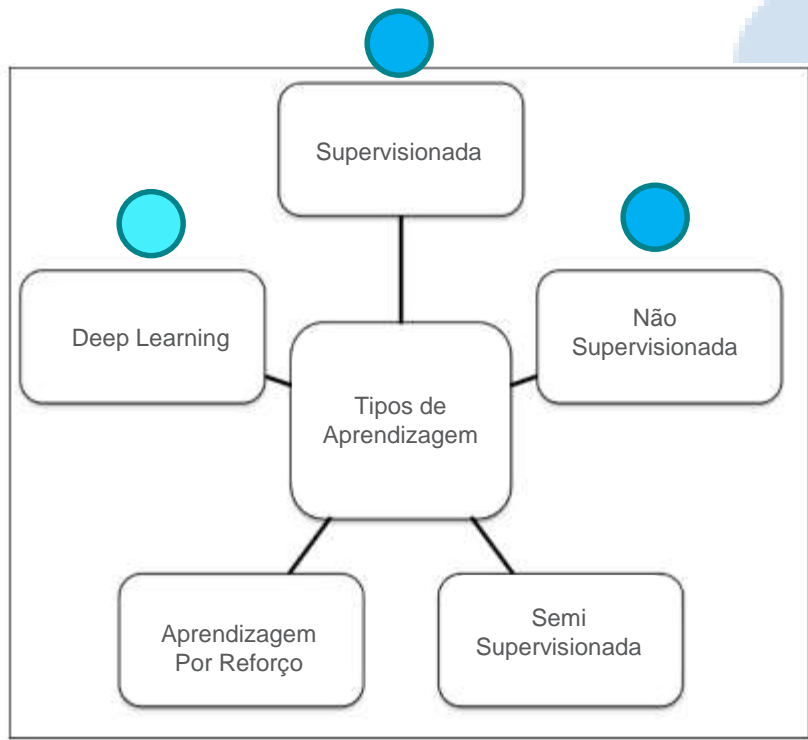






Machine Learning ou Aprendizado de Máquina é um método de análise de dados que automatiza o desenvolvimento de modelos analíticos. Usando algoritmos que aprendem interativamente a partir de dados, o aprendizado de máquinas permite que os computadores encontrem insights ocultos sem serem explicitamente programados para procurar algo específico.





Tipos de Aprendizagem






Mas se as máquinas estão aprendendo a aprender, isso significa que elas estão ficando inteligentes?





REPRODUZIR A

INTELIGÊNCIA HUMANA





Inteligência Artificial



Inteligência

Dotado de inteligência, capaz de compreender, esperto, habilidoso





Inteligência

Faculdade de conhecer, de aprender, de
conceber, de compreender:
a inteligência distingue o homem do animal





Inteligência Artificial

Conjunto de teorias e de técnicas empregadas com a finalidade de desenvolver máquinas capazes de simular a inteligência humana





Inteligência Artificial

A Inteligência Artificial é uma área de estudos da computação que se interessa pelo estudo e criação de sistemas que possam exibir um comportamento inteligente e realizar tarefas complexas com um nível de competência que é equivalente ou superior ao de um especialista humano





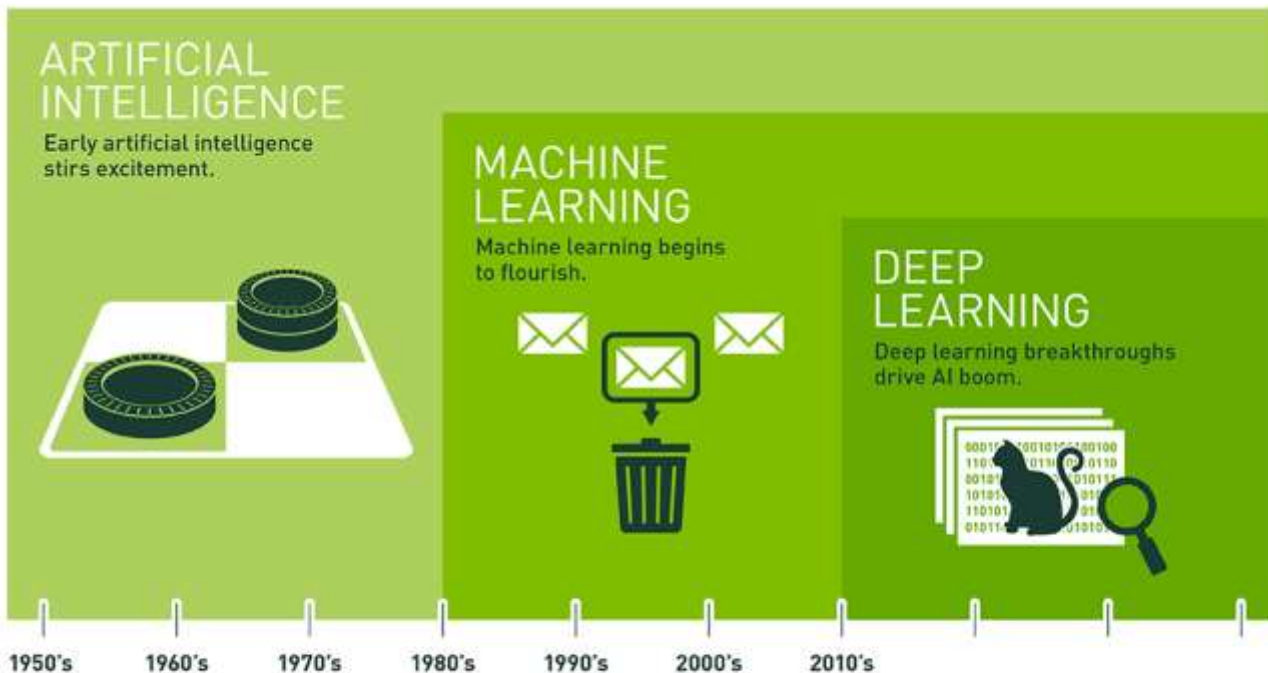


Inteligência Artificial

Estamos quase lá!







Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.



Aprendizado Não Supervisionado





Don't model the World; Model the Mind.





Data Science Academy

Por que Machine Learning Está
Transformando o Mundo?



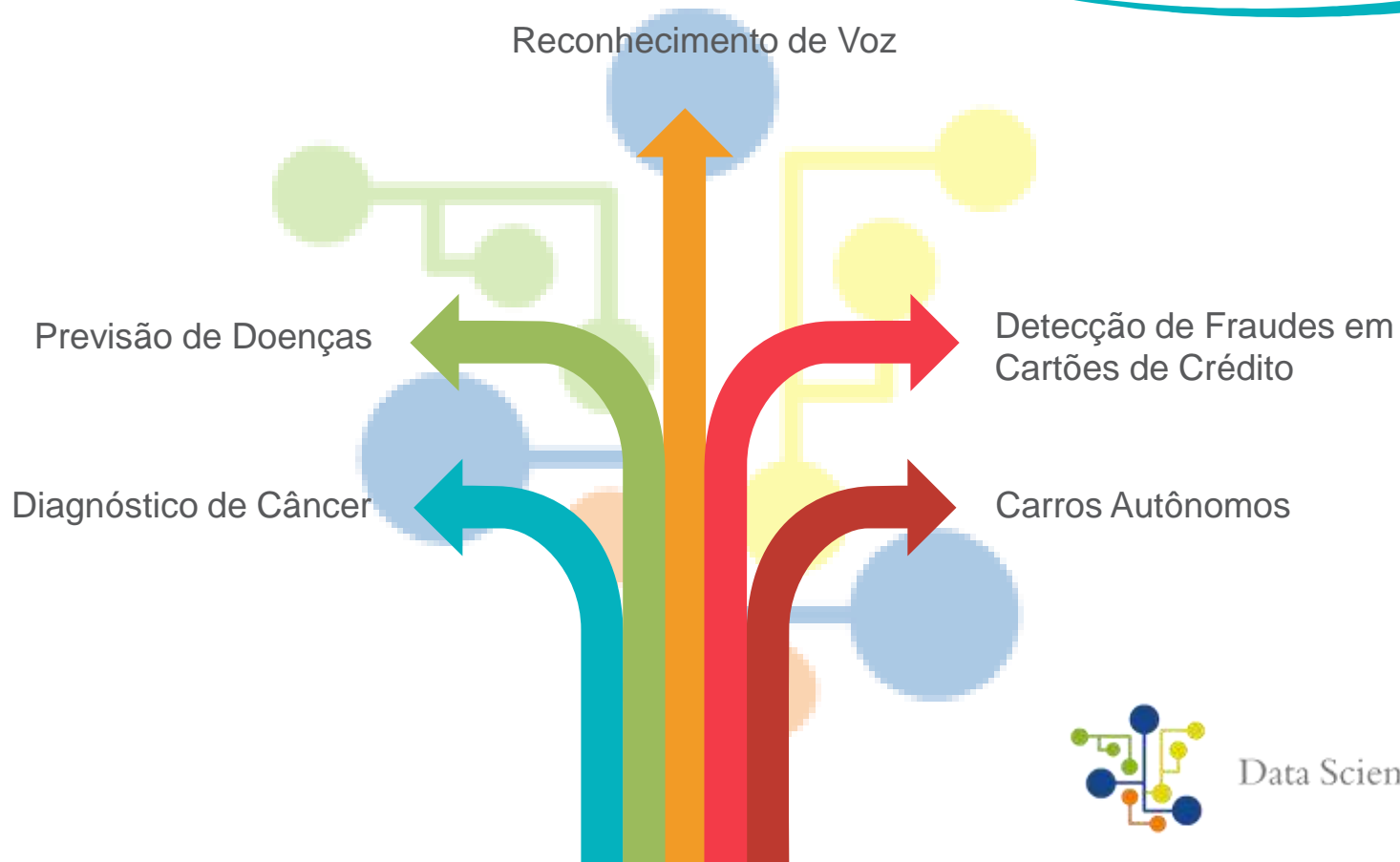
Data Science Academy



Algoritmos de aprendizagem de máquina, aprendem a induzir uma função ou hipótese capaz de resolver um problema a partir de dados que representam instâncias do problema a ser resolvido







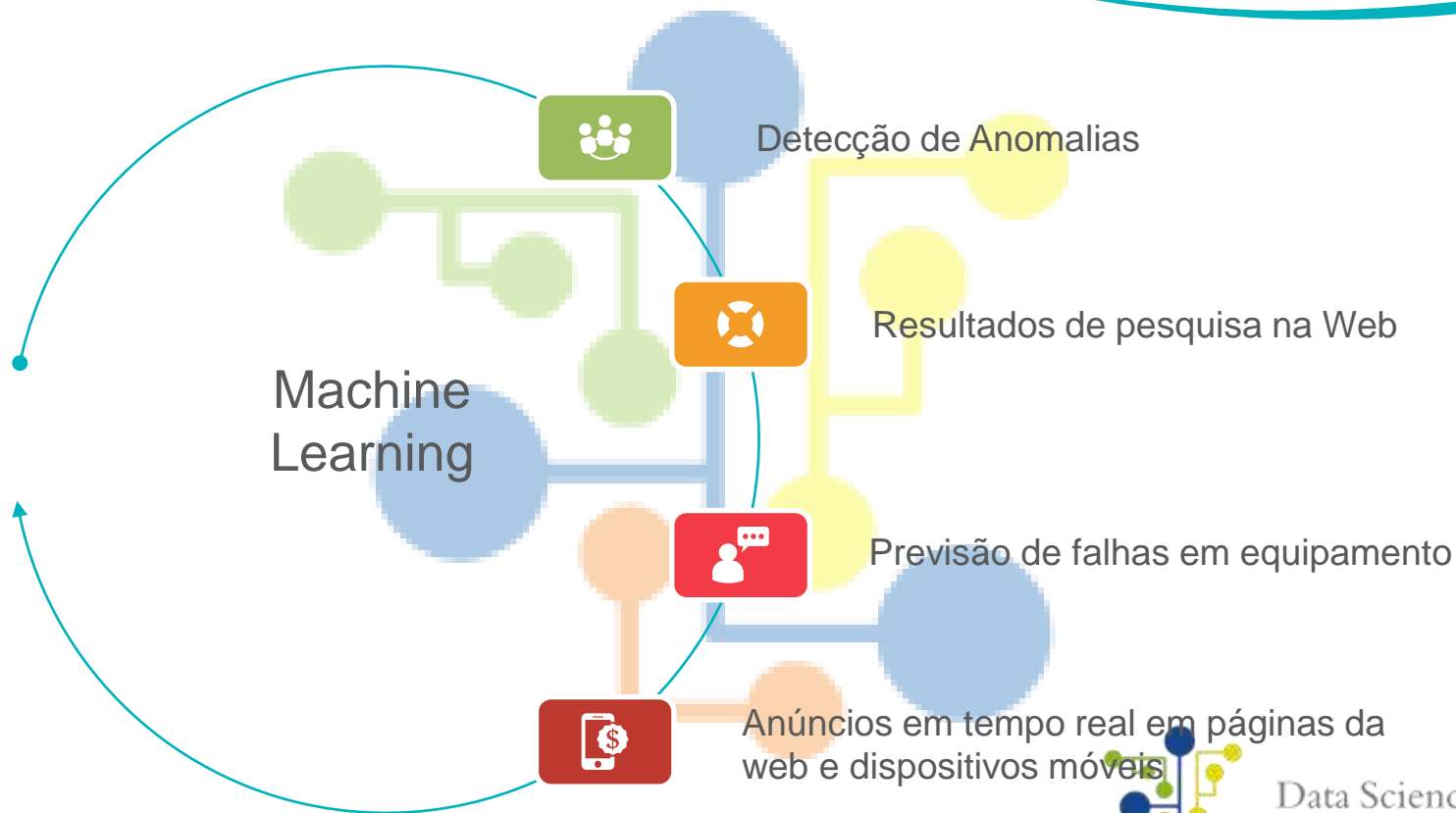


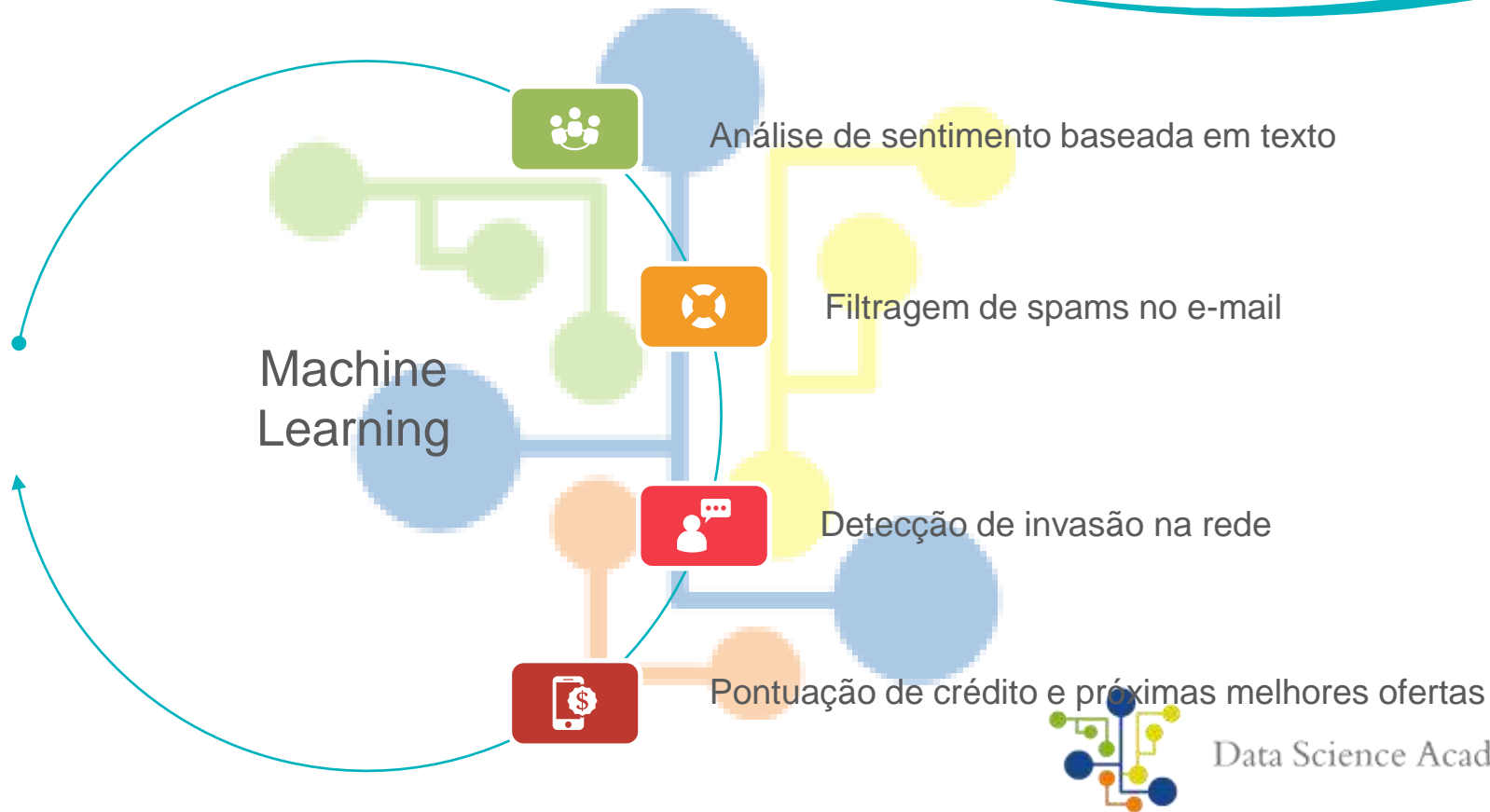


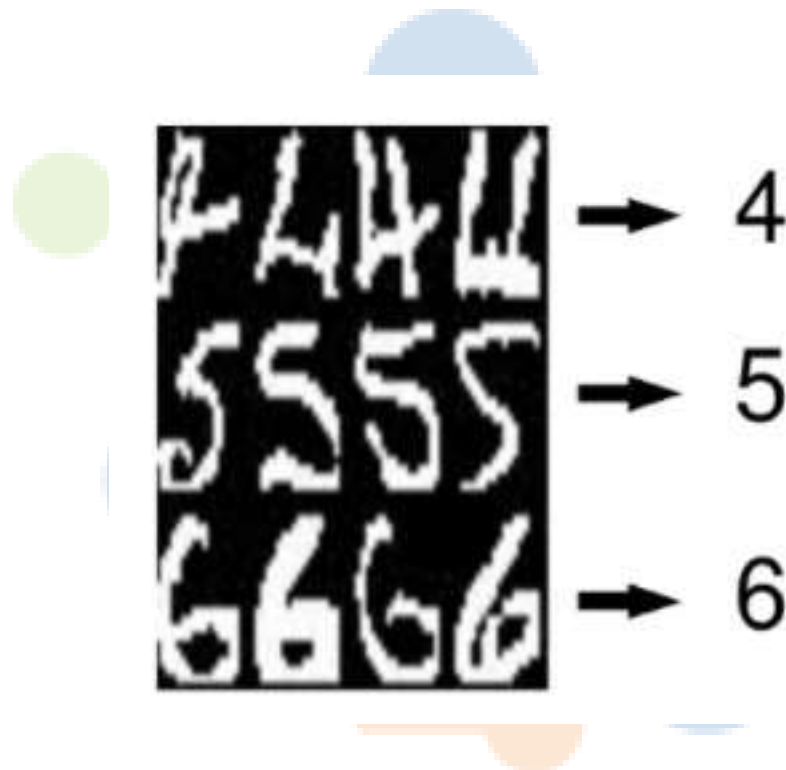
Algoritmos que Aprendem

As pessoas podem normalmente criar um ou dois bons modelos preditivos por semana; o aprendizado de máquina pode criar milhares de modelos por semana











Machine Learning não está transformando nosso mundo;





Machine Learning não está transformando nosso mundo;
Machine Learning já transformou o nosso mundo.





Data Science Academy

Que Ferramentas Usaremos Neste Curso?



Data Science Academy





theano



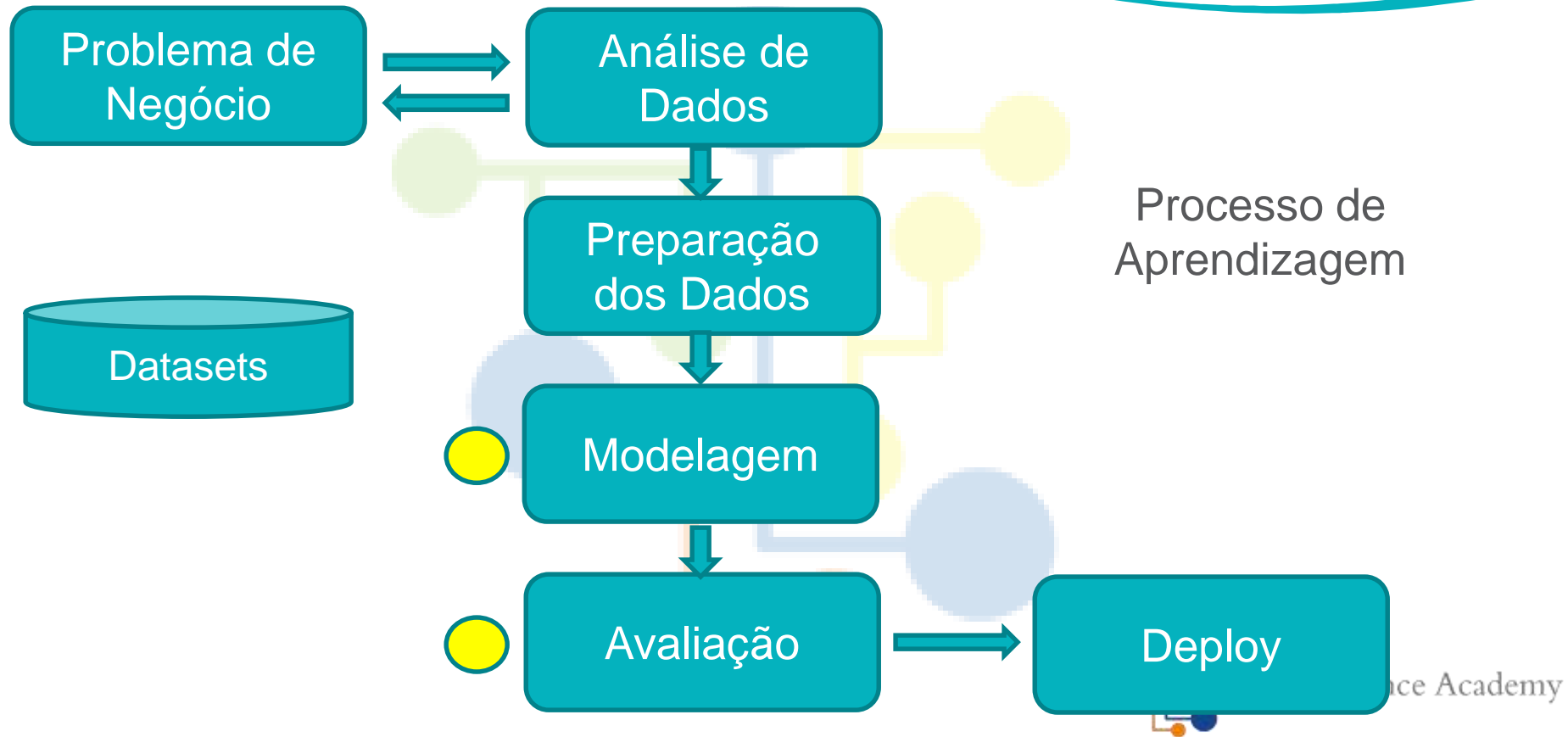


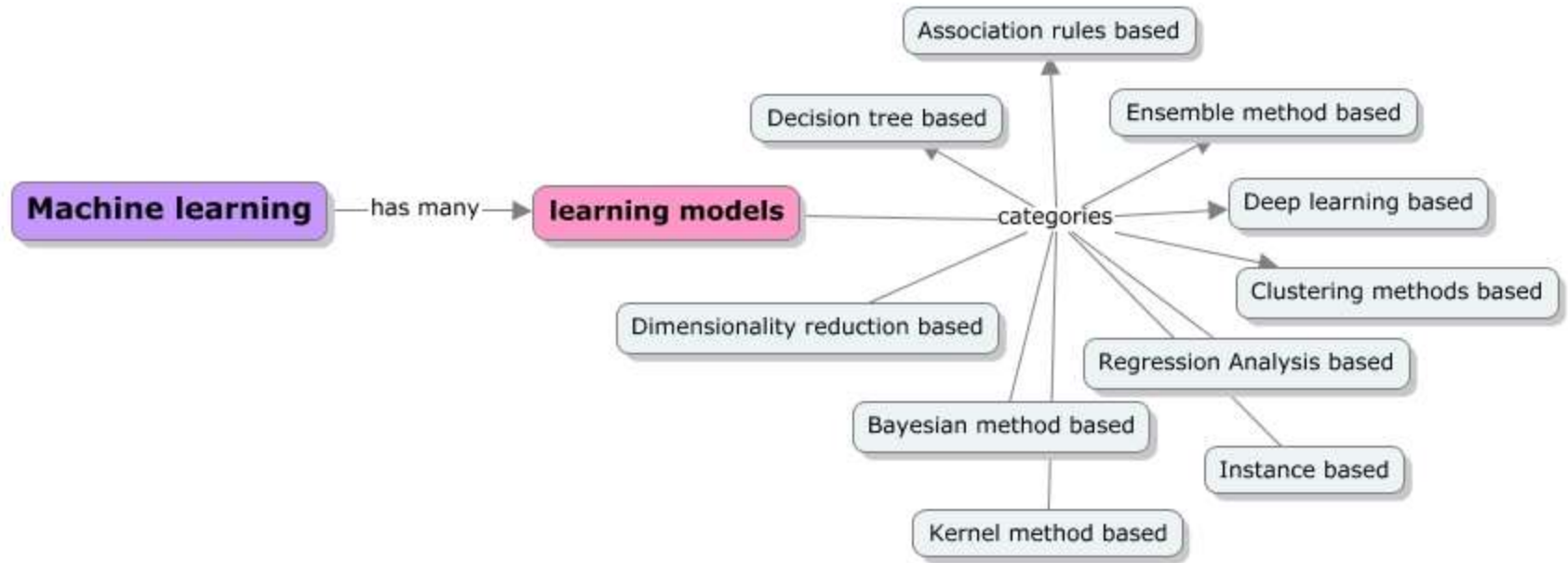
Data Science Academy

Processo de Aprendizagem



Data Science Academy







Data Science Academy

Tipos de Aprendizagem de Máquina



Data Science Academy



Aprendizagem
Supervisionada

Aprendizagem
Não
Supervisionada

Aprendizagem
Por Reforço





Data Science Academy

Aprendizagem Supervisionada

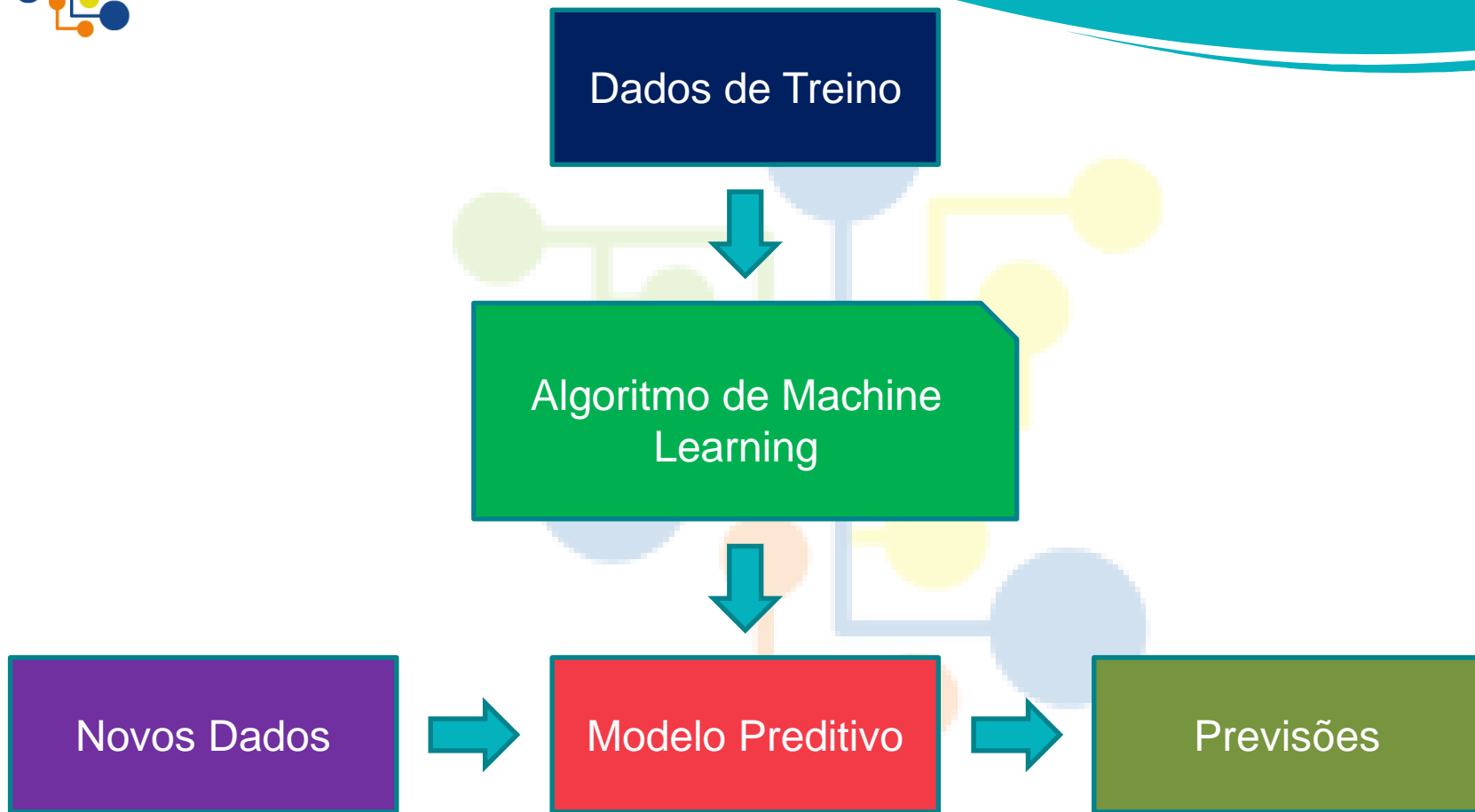


Data Science Academy



Aprendizagem Supervisionada







Os algoritmos de aprendizado supervisionado fazem previsões com base em um conjunto de exemplos





Aprendizagem
Supervisionada

Classificação

Regressão





**Aprendizagem
Supervisionada**



**Detecção de
Anomalias**





Aprendizagem Supervisionada

É o termo usado sempre que o programa é “treinado” sobre um conjunto de dados pré-definido





Data Science Academy

Aprendizagem Não Supervisionada



Data Science Academy



Aprendizagem Não Supervisionada





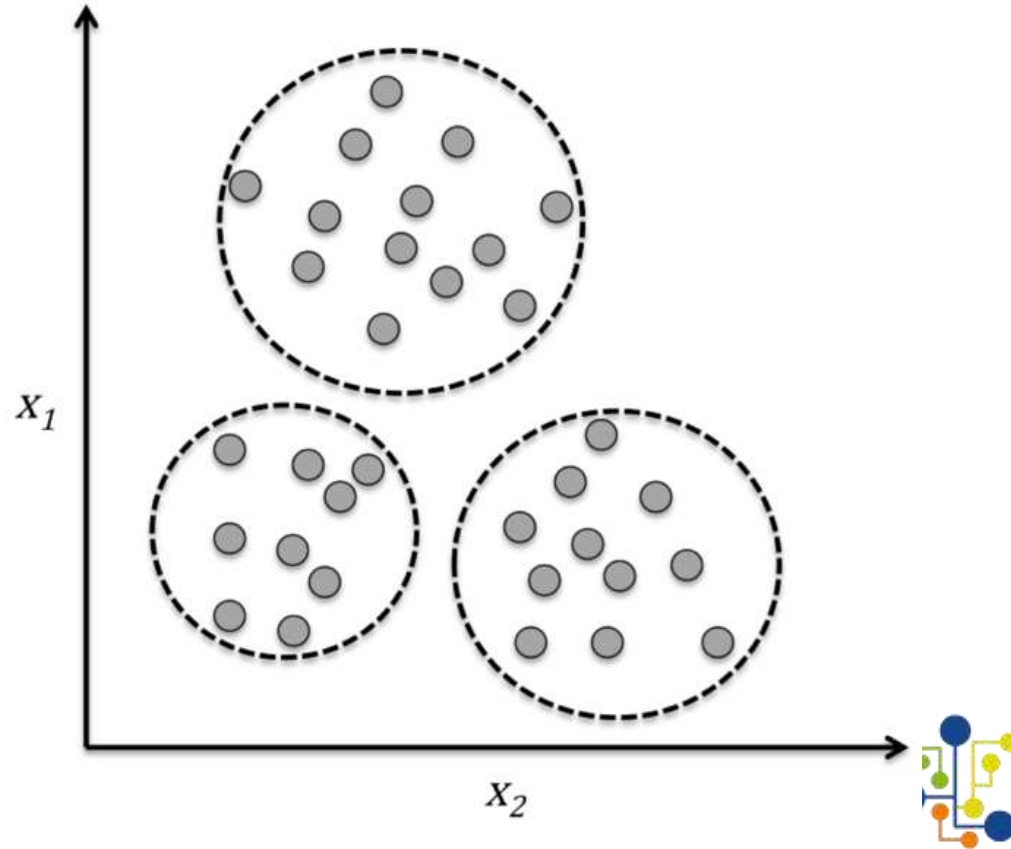
Este tipo de aprendizado, assemelha-se aos métodos que nós seres humanos usamos para descobrir se certos objetos ou eventos são da mesma classe





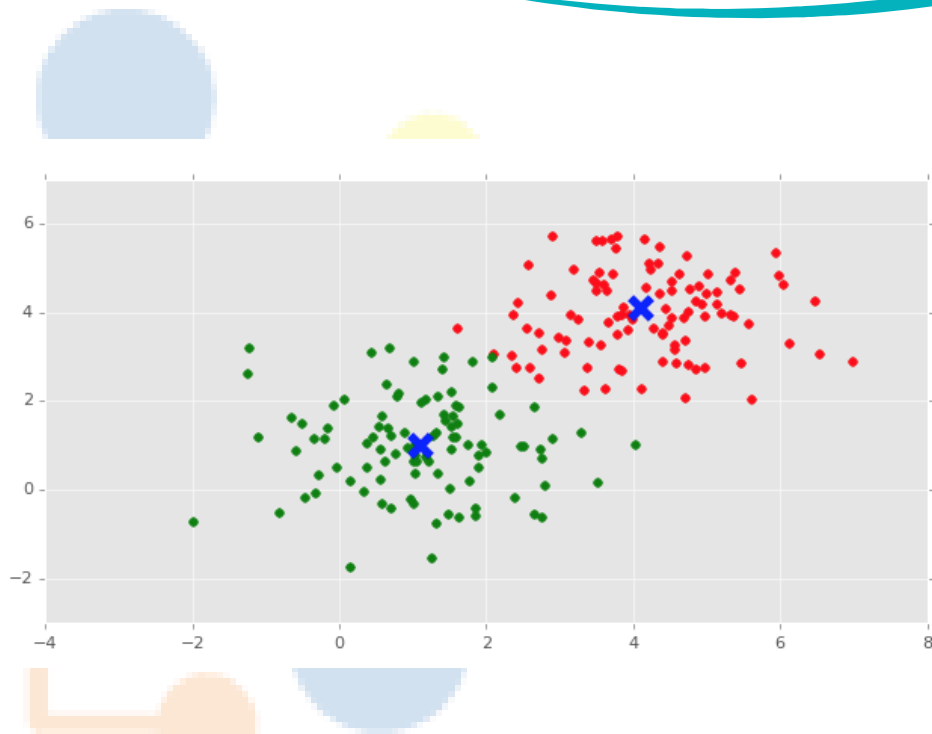
Alguns sistemas de recomendação que você encontra na internet sob a forma de automação de marketing são baseados neste tipo de aprendizagem







O objetivo de um algoritmo de aprendizado não supervisionado é organizar os dados de alguma forma ou descrever sua estrutura





Aprendizagem Não-Supervisionada

Termo usado quando um programa pode automaticamente encontrar padrões e relações em um conjunto de dados





Aprendizagem Não-Supervisionada

Os exemplos mais comuns são o K-Means, o Singular Value Decomposition (SVD) e o Principal Component Analysis (PCA)



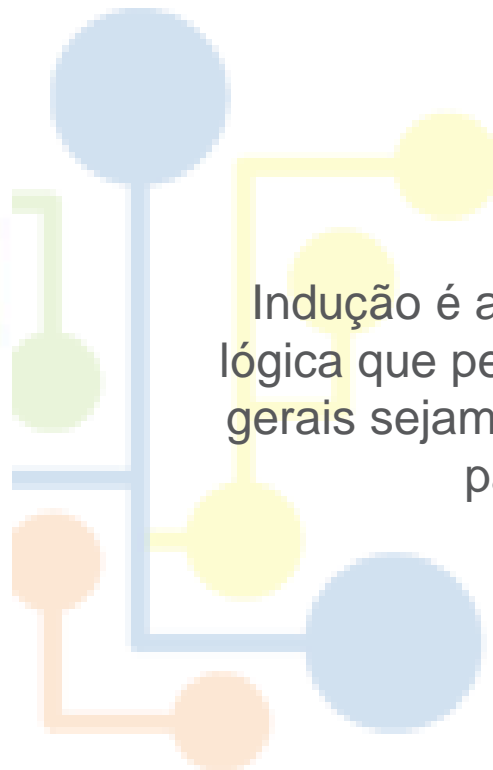


Data Science Academy

Aprendizado Indutivo



Data Science Academy



Indução é a forma de inferência lógica que permite que conclusões gerais sejam obtidas de exemplos particulares





O processo de indução é indispensável ao ser humano, pois é um dos principais meios de criar novos conhecimentos e prever eventos futuros









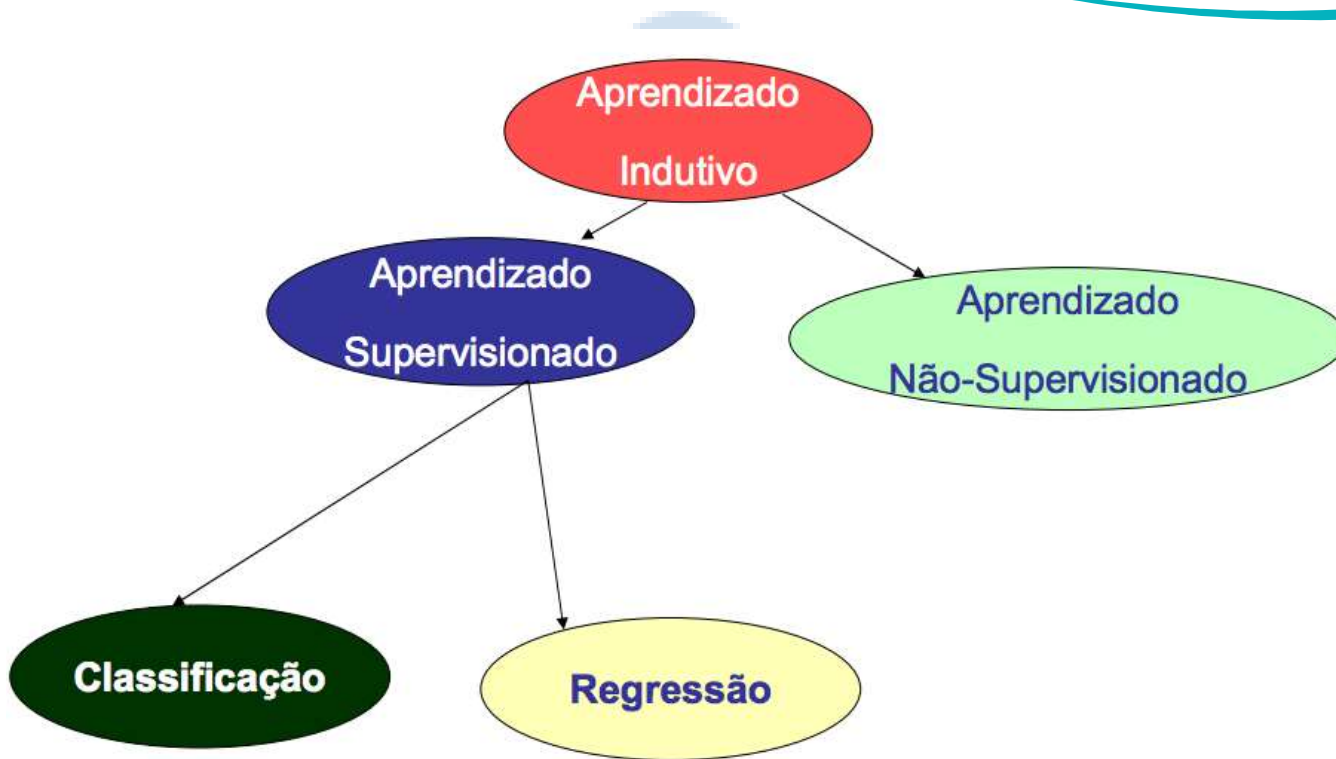
Aprendizado Supervisionado





Aprendizado Não
Supervisionado







Data Science Academy

Reinforcement Learning (Aprendizagem por Reforço)



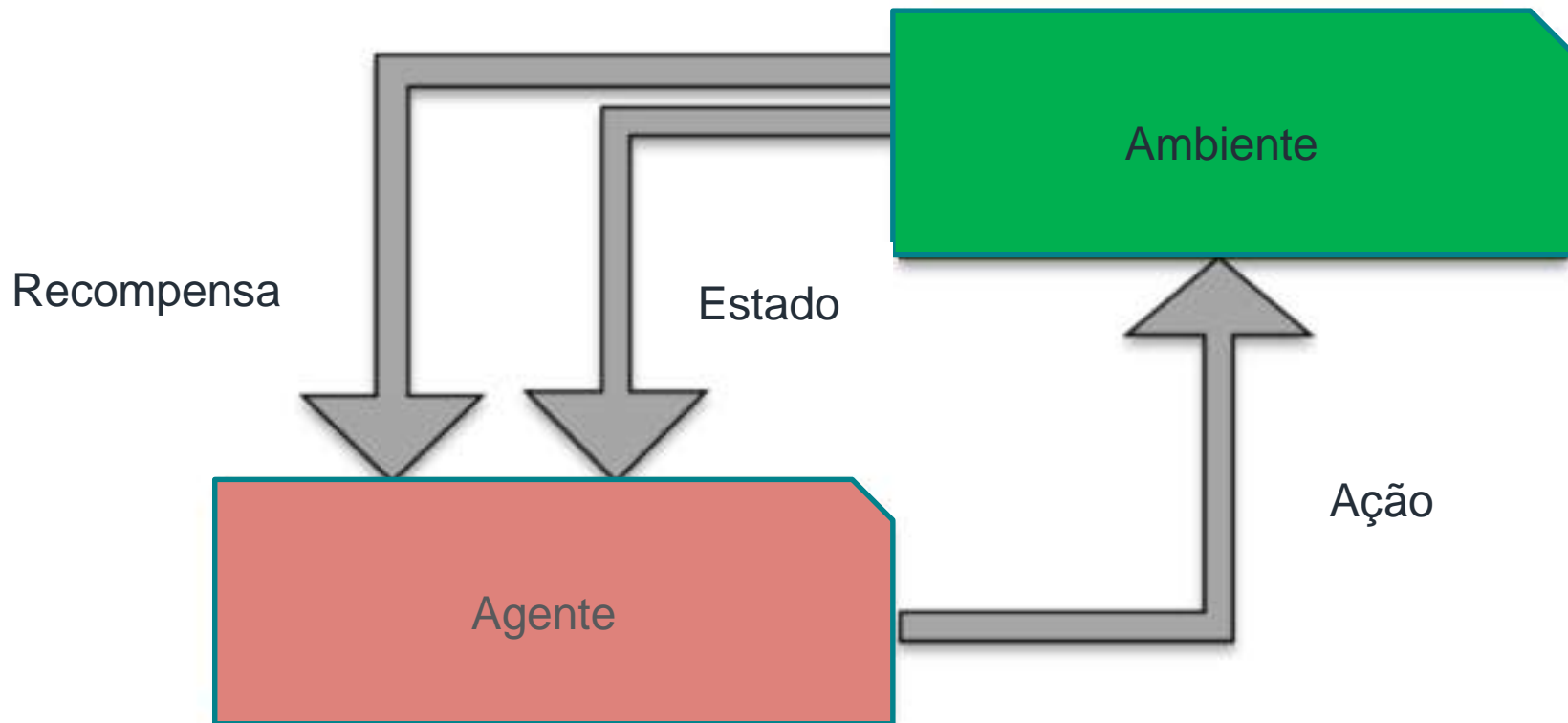
Data Science Academy



Reinforcement Learning

Similar ao que chamamos de aprender por tentativa e erro





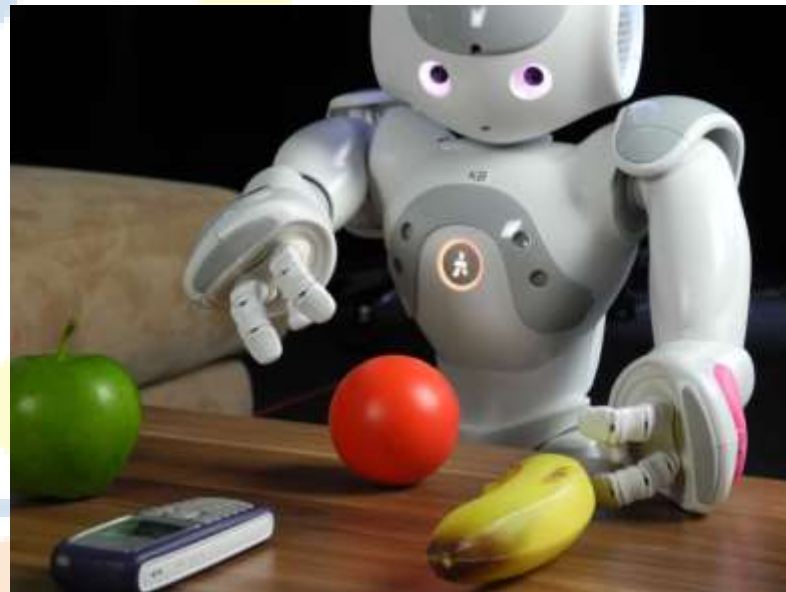


No aprendizado por reforço, o algoritmo escolhe uma ação em resposta a cada ponto de dados





O aprendizado por reforço é comum em robótica, em que o conjunto de leituras do sensor, em um ponto no tempo, é um ponto de dados e o algoritmo deve escolher a próxima ação do robô





O aprendizado por reforço é definido não caracterizando algoritmos de aprendizado, mas sim o problema a ser aprendido



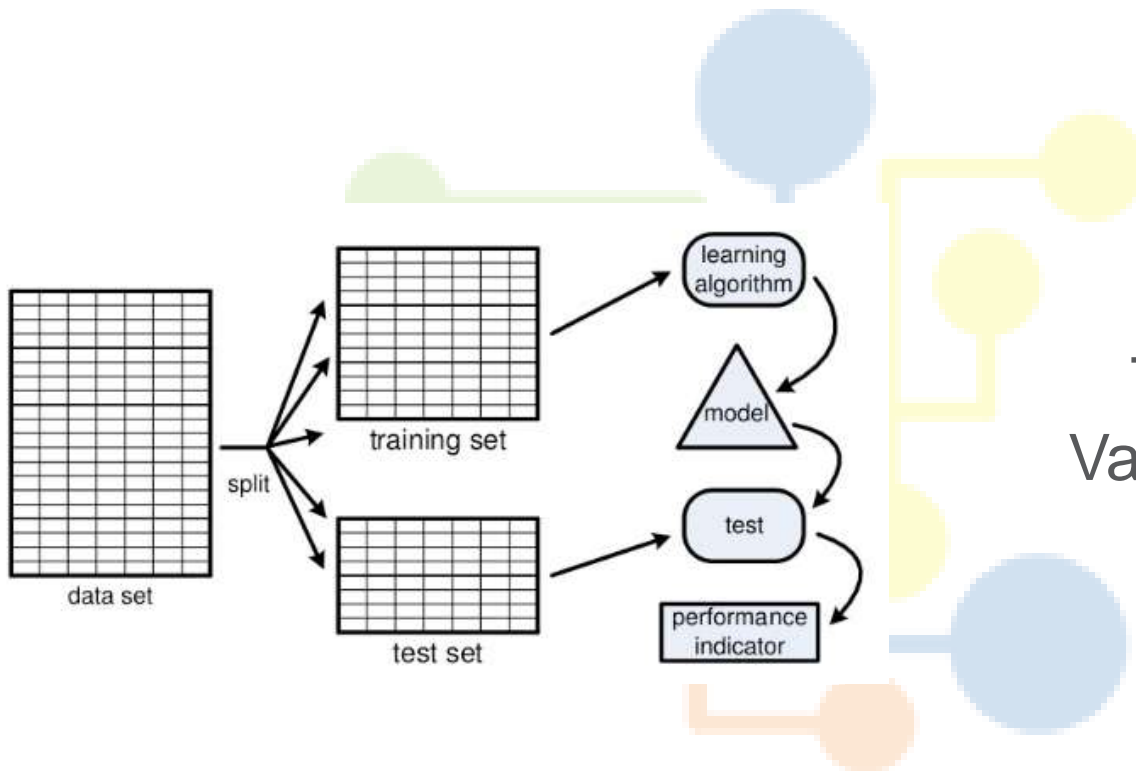


Data Science Academy

Treinamento, Validação e Teste



Data Science Academy



Treinamento, Validação e Teste





Treinamento, Validação e Teste

75 a 70% - dados de treino
25 a 30% - dados de teste





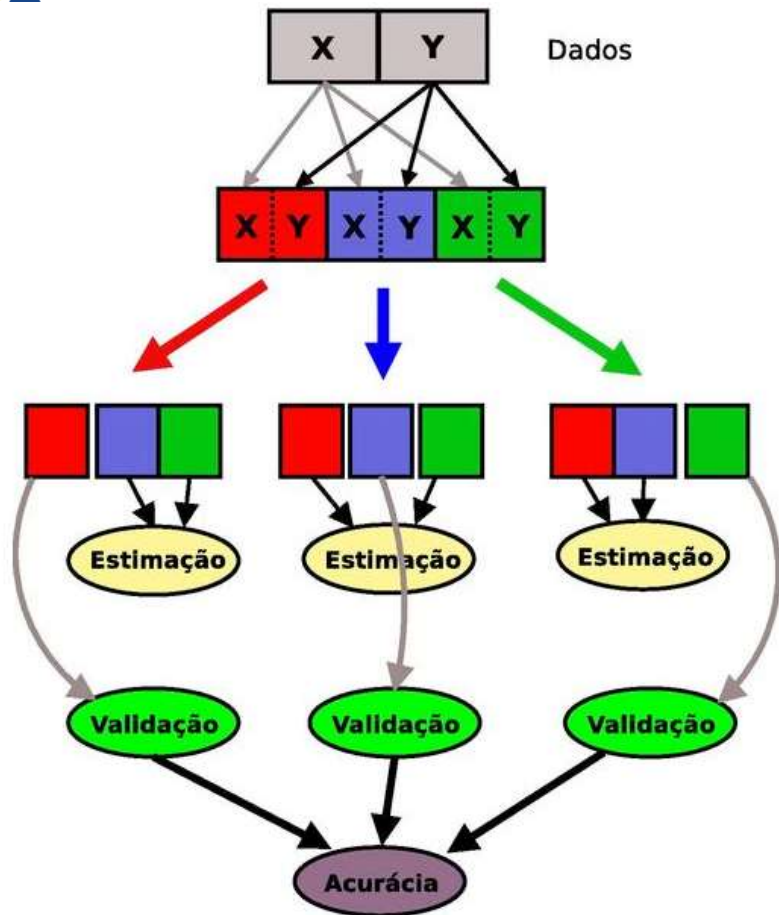
Treinamento, Validação e Teste

75 a 70% - dados de treino

20% - dados de validação

10% - dados de teste





Treinamento,
Validação e Teste





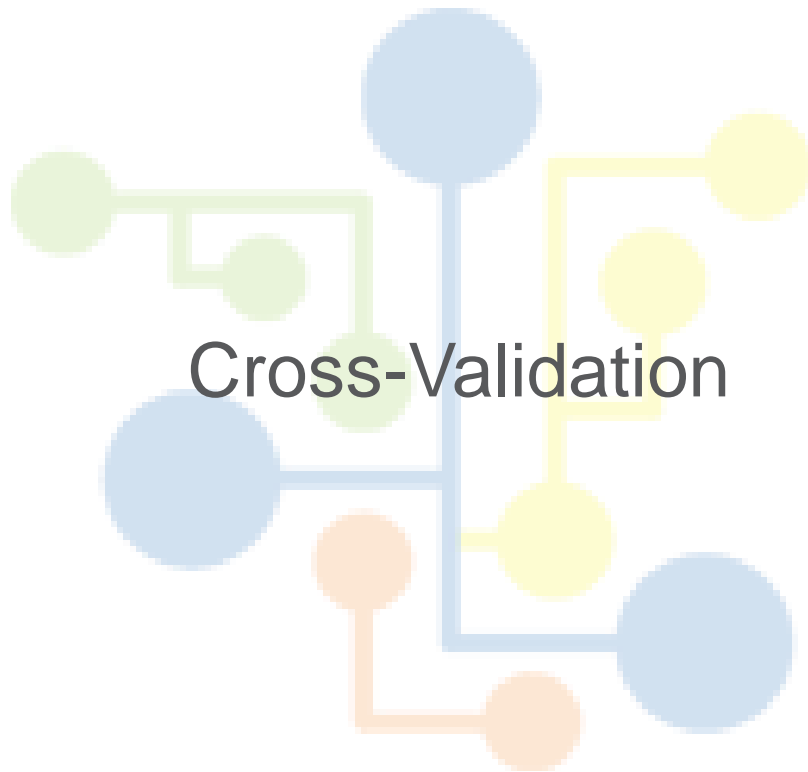
Treinamento, Validação e
Teste

$n > 10.000$



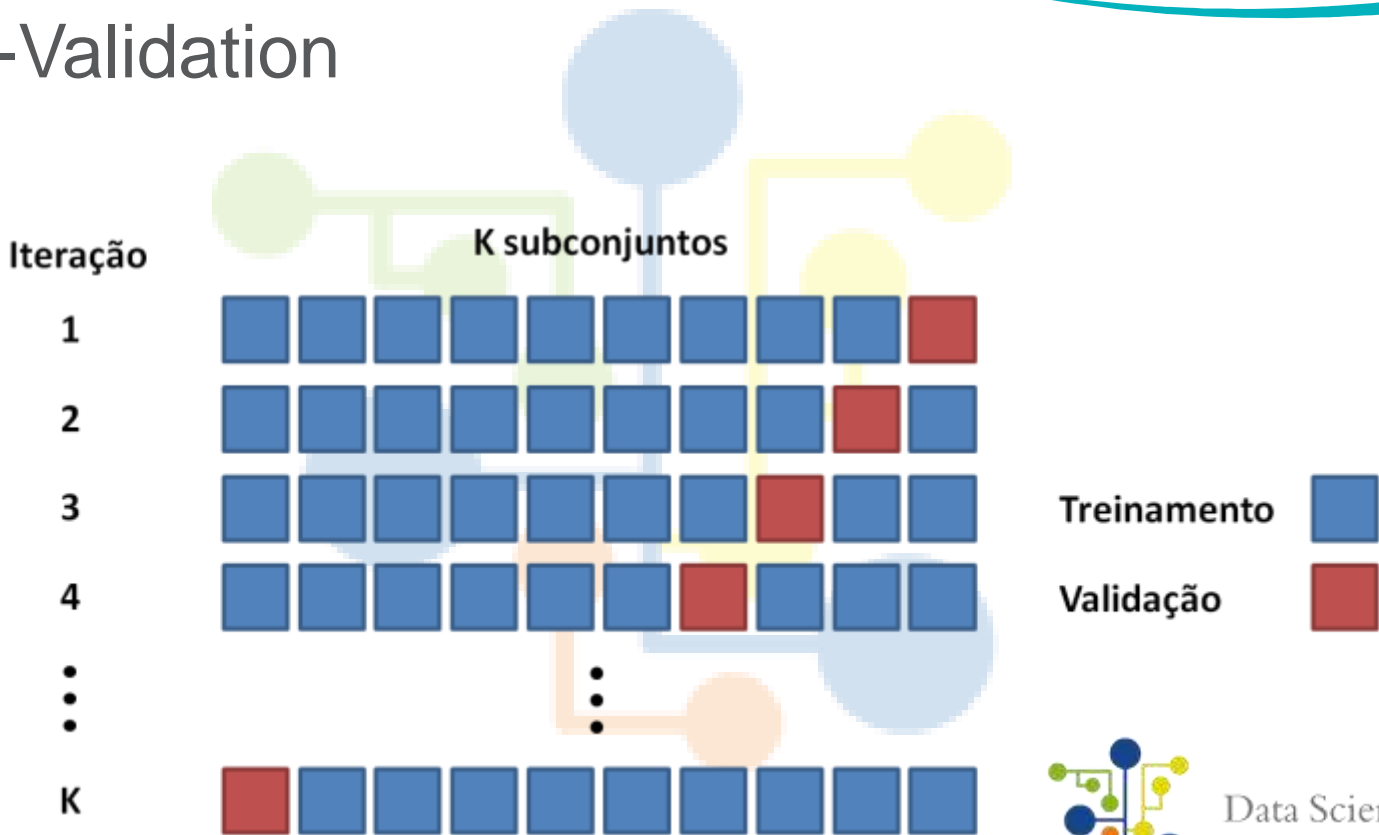


Cross-Validation





Cross-Validation



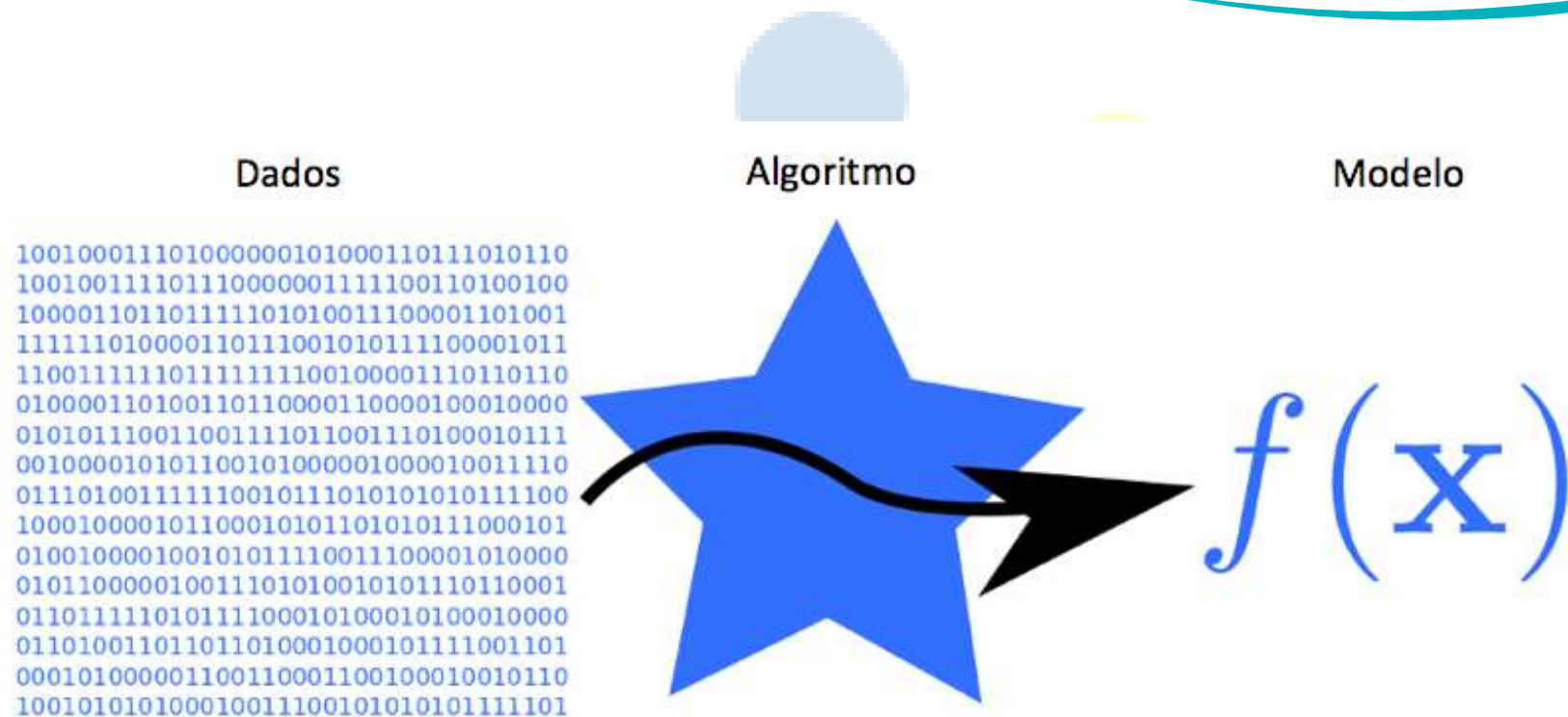


Data Science Academy

O que é um Modelo?



Data Science Academy





Modelo

Observações



Dados

Distance	Time
4.9m	1s
19.6m	2s
44.1m	3s
78.5m	4s



Modelo

$$g = 9.8m/s^2$$





Modelo



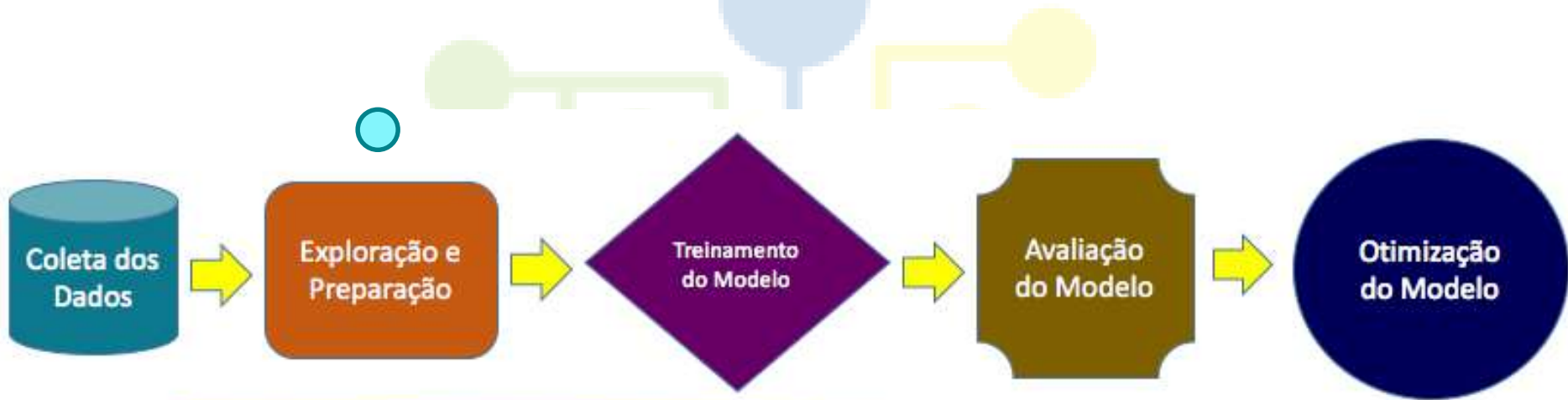


Modelo





Modelo





Modelo





Modelo

O processo de "fitting" um modelo a um dataset é chamado de treinamento do modelo





Modelo



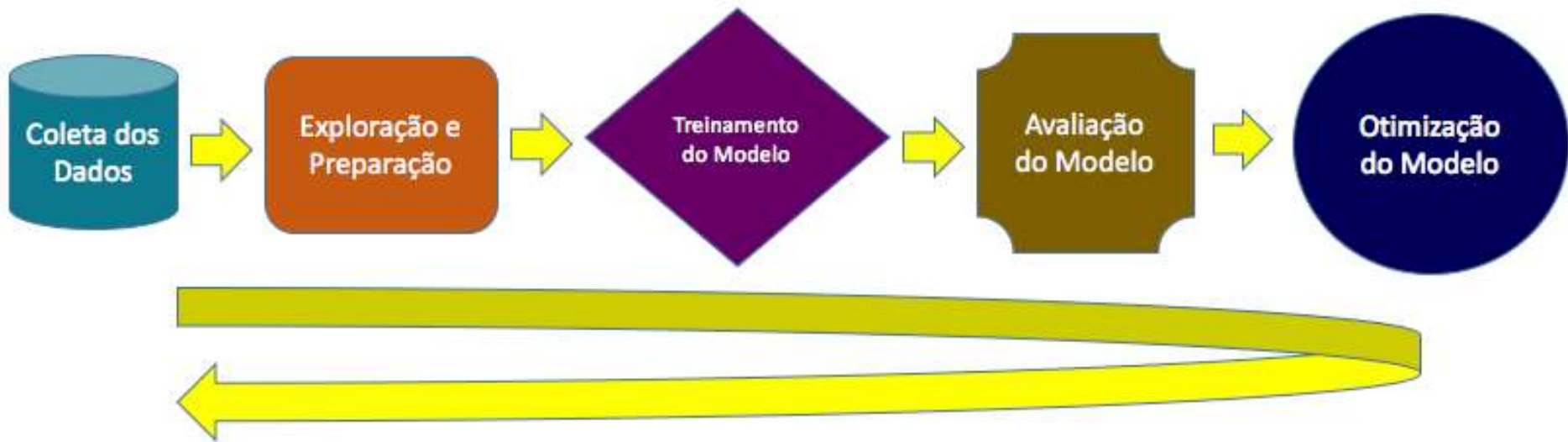


Modelo





Modelo





Seu trabalho como Cientista de
Dados é buscar sempre o
melhor modelo possível para
suas previsões





O modelo pode ser implantado para resolver o problema de negócio para o qual ele foi desenvolvido





Lembre-se: um modelo de Machine Learning será usado para resolver um problema específico





Não caia na tentação de querer
aplicar seu modelo a tudo que você
vê pela frente



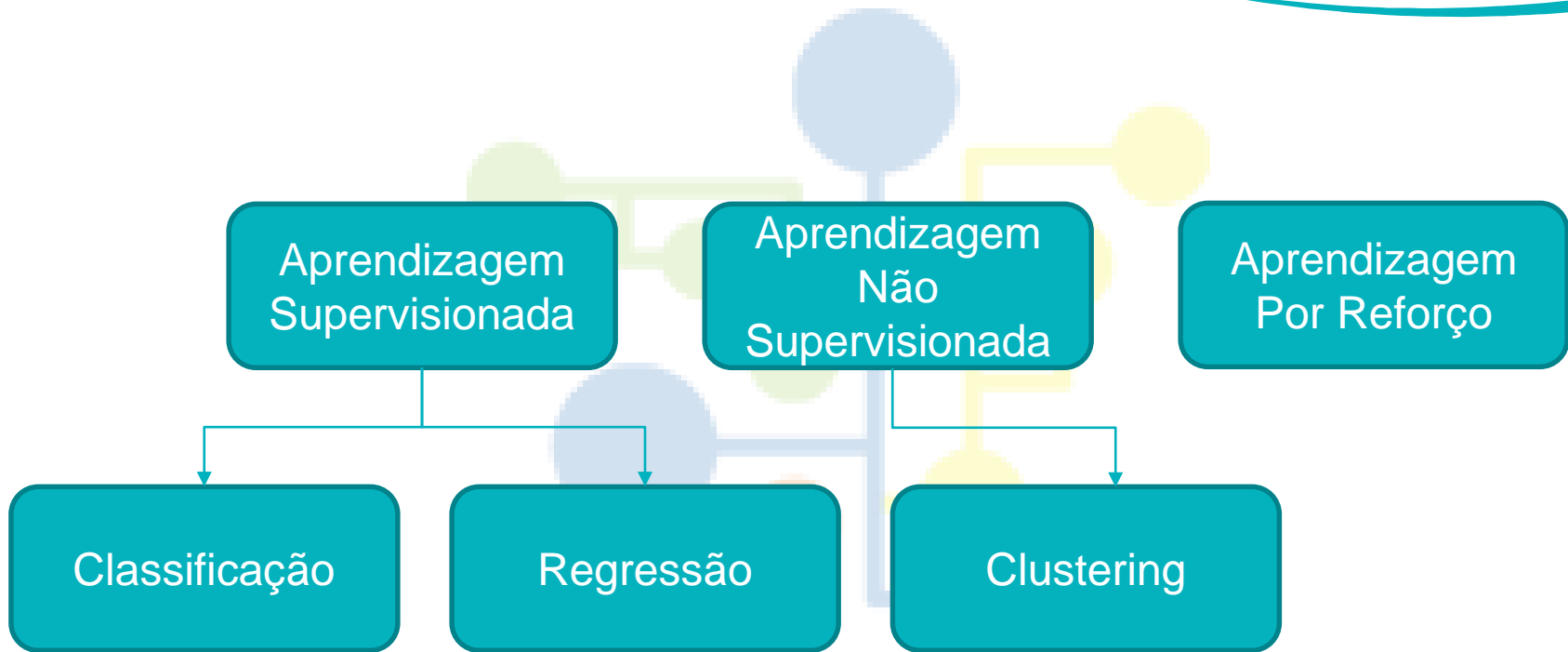


Data Science Academy

Classificação



Data Science Academy





Podemos representar a realidade e toda sua complexidade através de funções matemáticas





Classificação

É o processo de identificar a qual conjunto de categorias uma nova observação pertence, com base em um conjunto de dados de treino contendo observações (ou instâncias) cuja associação é conhecida

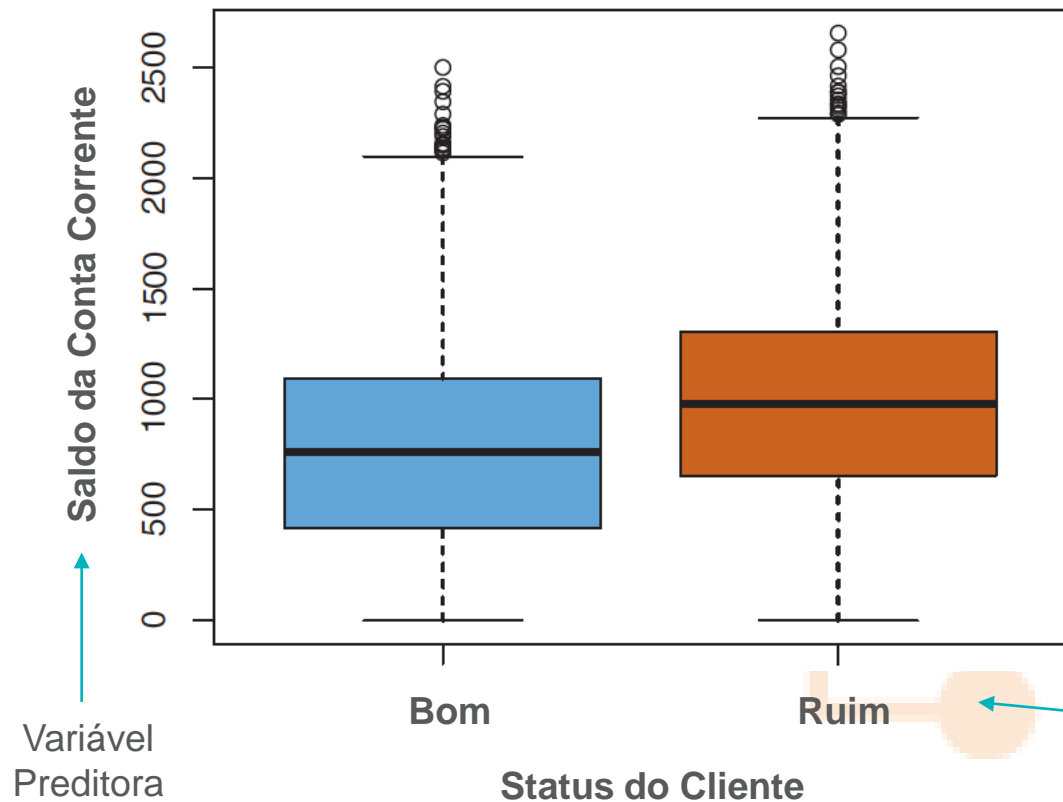




Classificação

Exemplo: determinar o diagnóstico de uma doença em um paciente, observando as características similares em outros grupos de pacientes





Classificação

Variável Target
Pode assumir os valores:
Bom ou Ruim



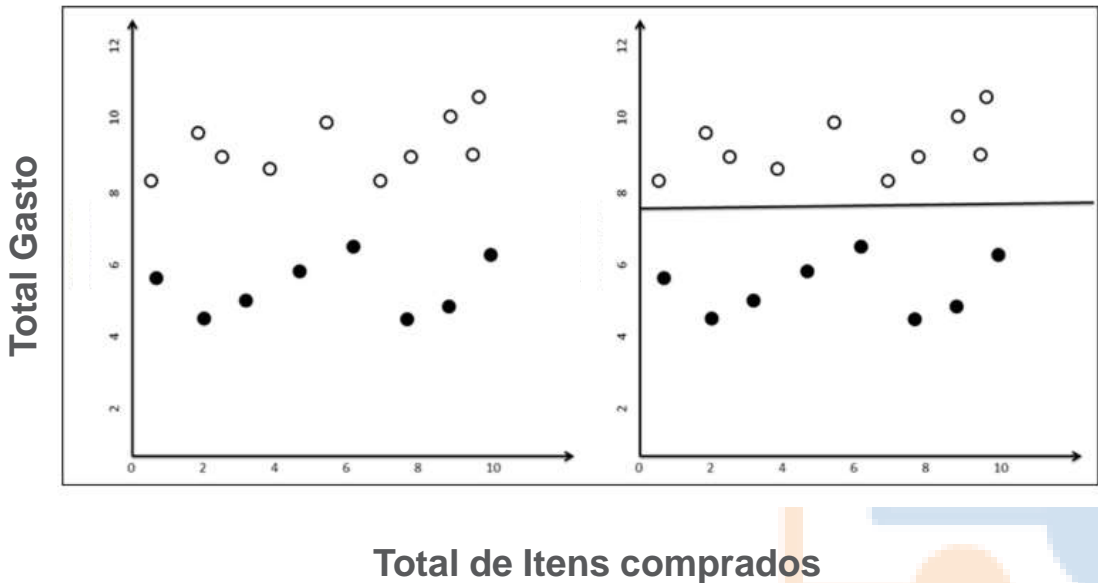
Classificação





Classificação

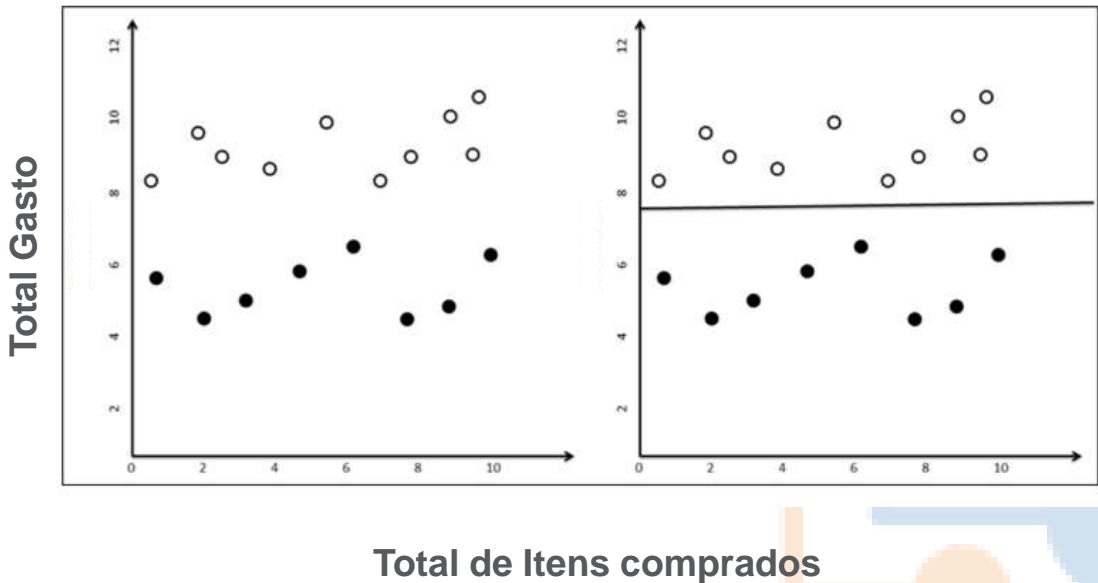




Classificação

- Sim (investir em pré-venda)
- Não (não investir em pré-venda)

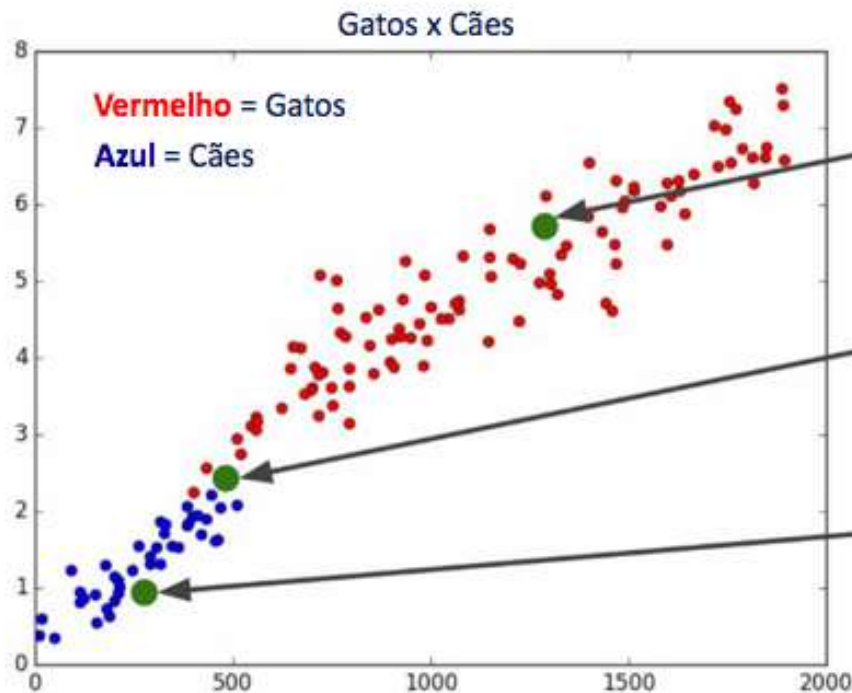


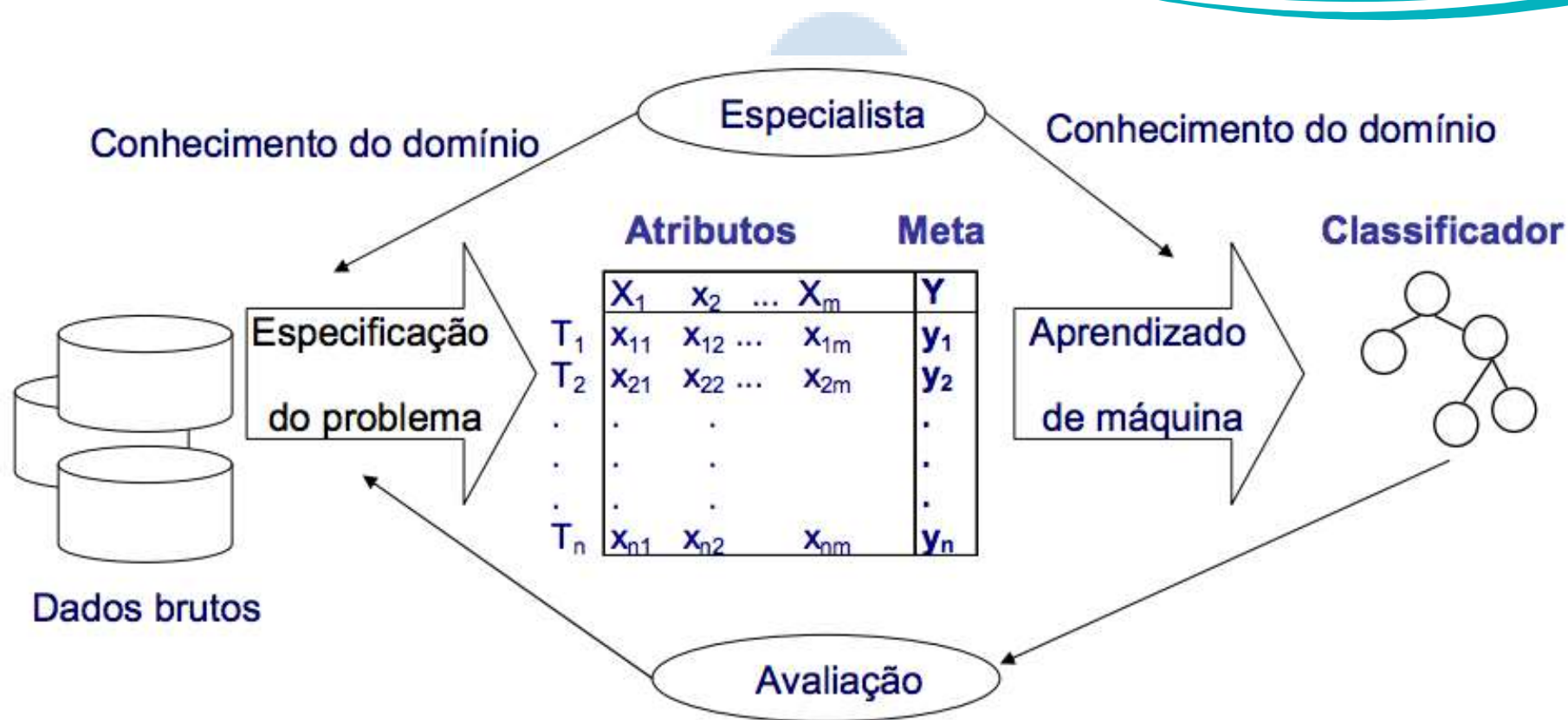


Classificação

- Sim (investir em pré-venda)
- Não (não investir em pré-venda)









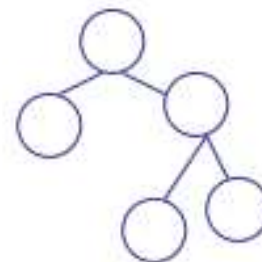
Atributos

Classe

	X_1	x_2	...	X_m	Y
T_1	x_{11}	x_{12}	...	x_{1m}	y_1
T_2	x_{21}	x_{22}	...	x_{2m}	y_2
.	.	.			.
.	.	.			.
.	.	.			.
T_n	x_{n1}	x_{n2}		x_{nm}	y_n



Classificador



Hipótese

Descrição de conceito



Bias

Qualquer preferência de uma hipótese sobre a outra



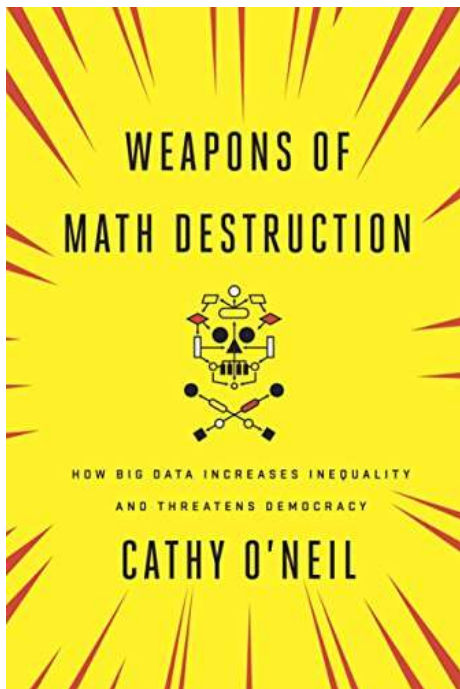


Data Science Academy

The Dark Side of Big Data



Data Science Academy



Recomendo

Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy





Data Science Academy

Regressão



Data Science Academy



Regressão

Um estudo de regressão busca, essencialmente, associar uma variável Y (denominada variável resposta ou variável dependente) a uma outra variável X (denominada variável explanatória ou variável independente)





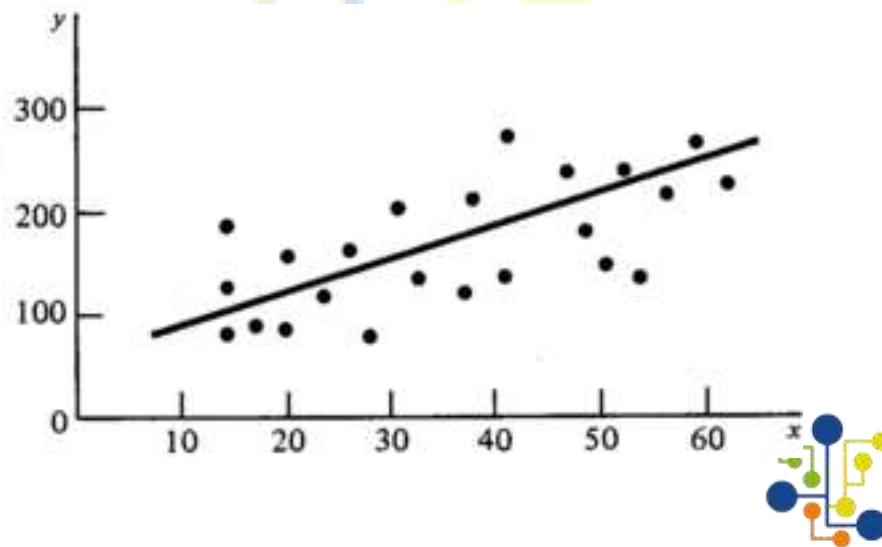
Como a Regressão pode ser usada?

- Investigação Científica
- Relações Causais
- Indentificação de Padrões





Compreendendo a Regressão



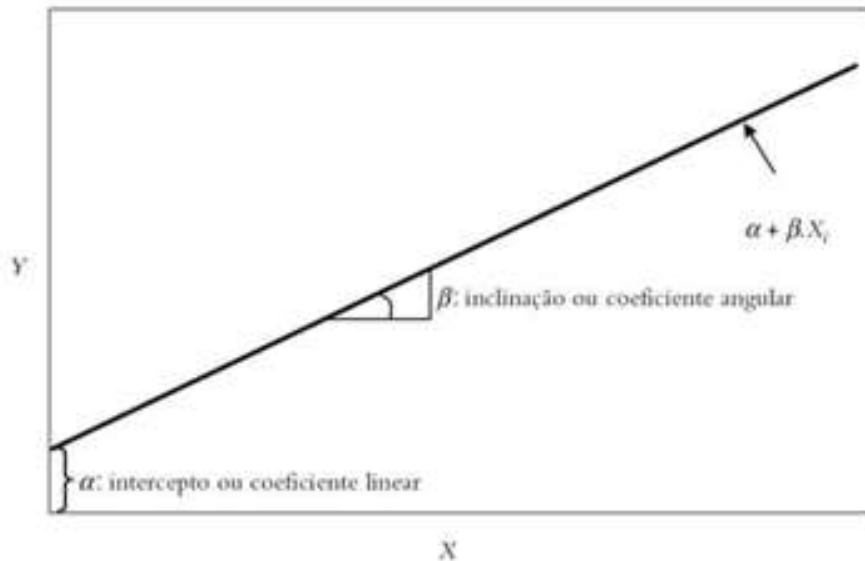


Compreendendo a Regressão

$$\hat{y} = a + bx$$

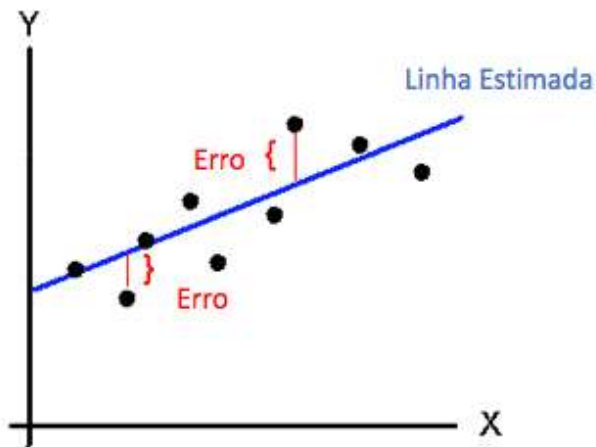
Onde:

- \hat{y} = valor previsto de y dado um valor para x
- x = variável independente
- a = ponto onde a linha intercepta o eixo y
- b = inclinação da linha reta





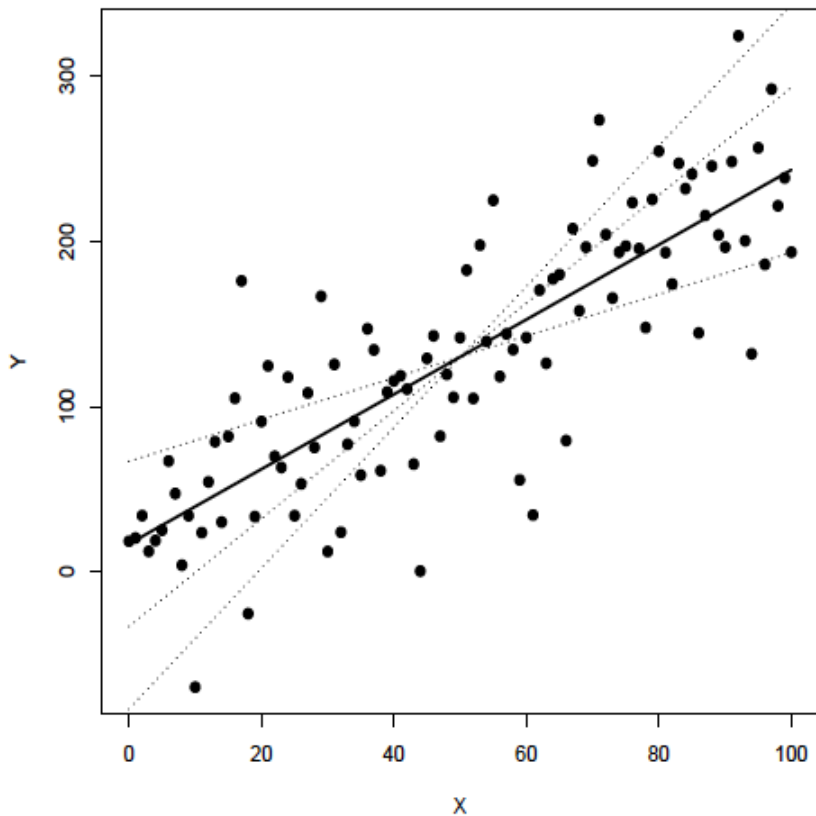
Estimativa dos Mínimos Quadrados





Deve-se determinar α e β de modo que a somatória dos quadrados dos resíduos seja a menor possível (método de Mínimos Quadrados Ordinários - MQO, ou, em inglês, Ordinary Least Squares - OLS)





Os coeficientes dessa reta
podem ser estimados pelo
Método dos Quadrados Mínimos





Correlação

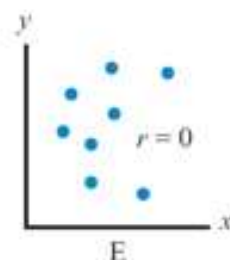
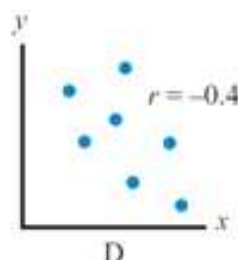
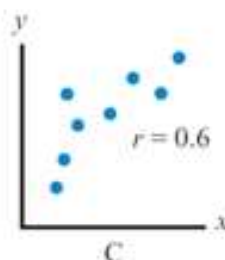
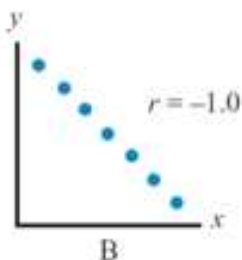
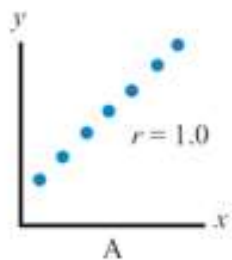


Gráfico A ($r = 1.0$): correlação positiva perfeita entre x e y

Gráfico B ($r = -1.0$): correlação negativa perfeita entre x e y

Gráfico C ($r = 0.6$): relação positiva moderada: y tende a aumentar se x aumenta, mas não necessariamente na mesma taxa observada no Gráfico A

Gráfico D ($r = -0.4$): relação negativa fraca: o coeficiente de correlação é próximo de zero ou negativo: y tende a diminuir se x aumenta

Gráfico E ($r = 0$): Sem relação entre x e y



Os valores de r variam entre **-1.0** (uma forte relação negativa) até **+1.0**, uma forte relação positiva.





Correlação Não Implica Causalidade





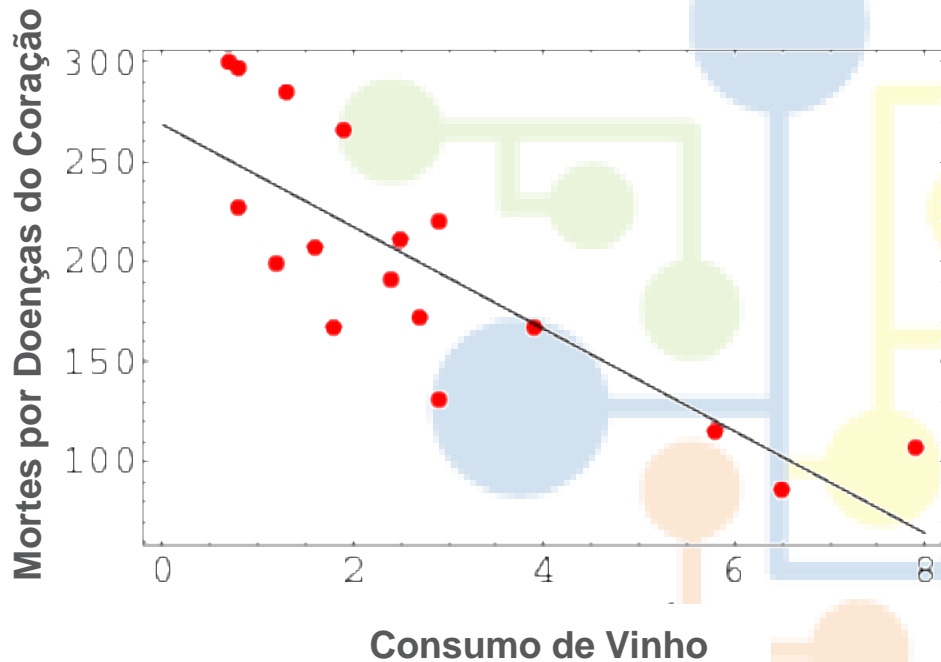
Só porque (A) acontece juntamente com (B)
não significa que (A) causa (B)





Regressão



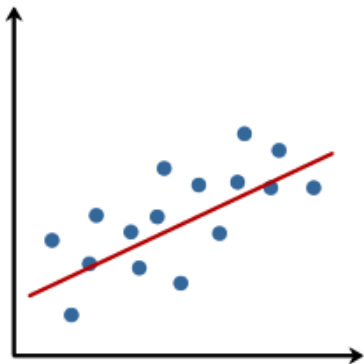


Regressão

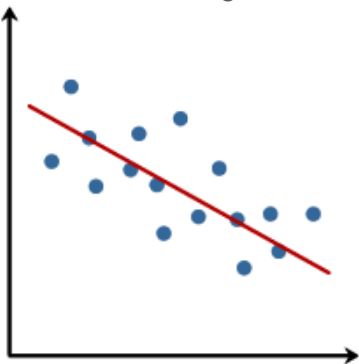




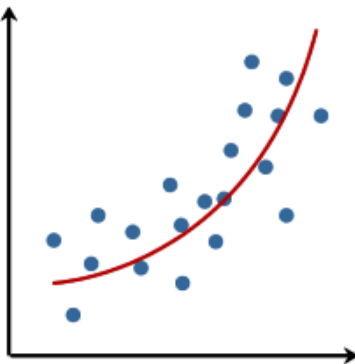
Relacionamento
Linear Positivo



Relacionamento
Linear Negativo



Sem Relacionamento
Linear



Regressão







Data Science Academy

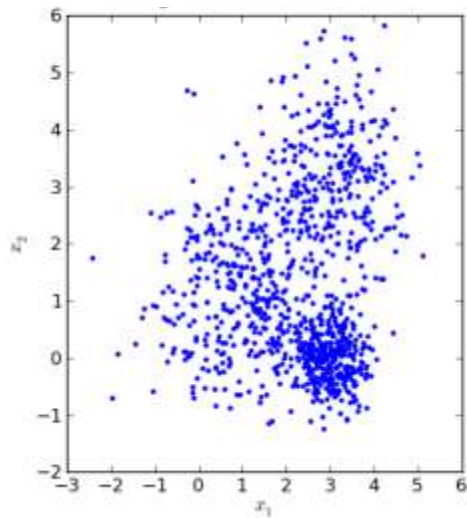
Clustering



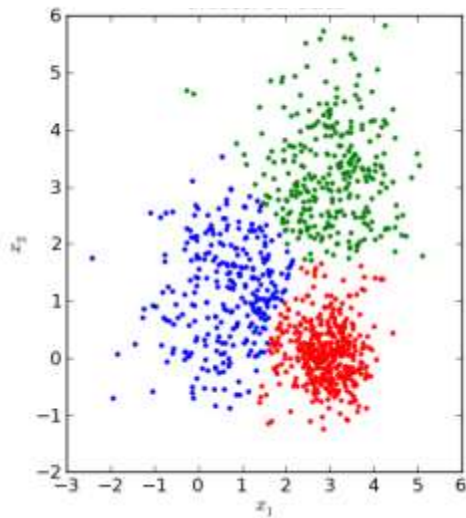
Data Science Academy



Antes da Clusterização



Depois da Clusterização



Clustering





Algoritmos de Aprendizagem Não Supervisionada

Categoria	Algoritmo
Algoritmos Baseados em Centroides	K-means, Gaussian Mixture Model, Fuzzy c-mean
Algoritmos Baseados em Conectividade	Algoritmos hierárquicos
Algoritmos Baseados em Densidade	DBSCAN, Optics
Probabilísticos	LDA
Redução de Dimensionalidade	tSNE, PCA, KPCA
Redes Neurais / Deep Learning	Autoencoders



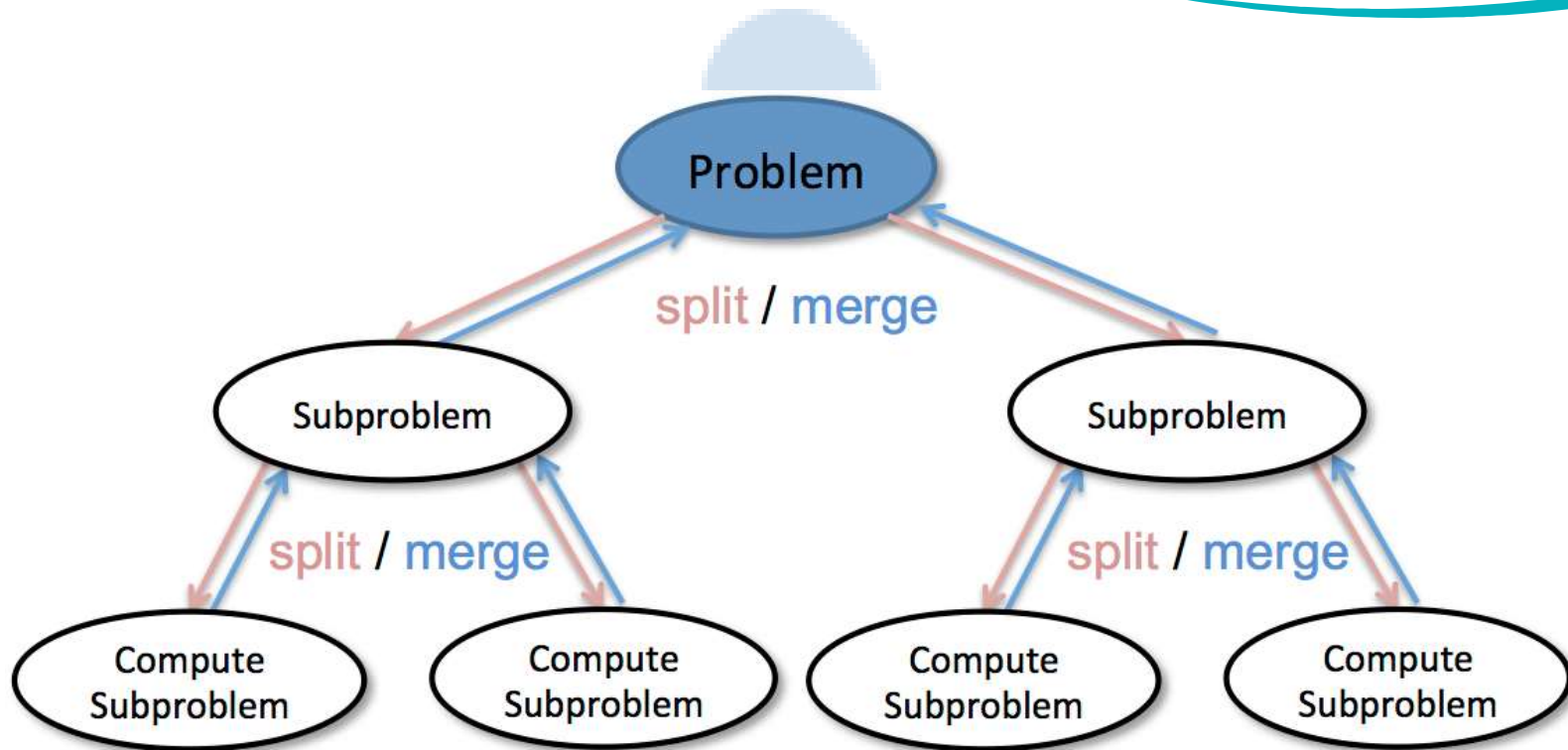


Data Science Academy

Machine Learning é Dividir Para Conquistar



Data Science Academy







Mineração de
Dados

Aprendizagem
de Máquina

Aprendizagem
Profunda

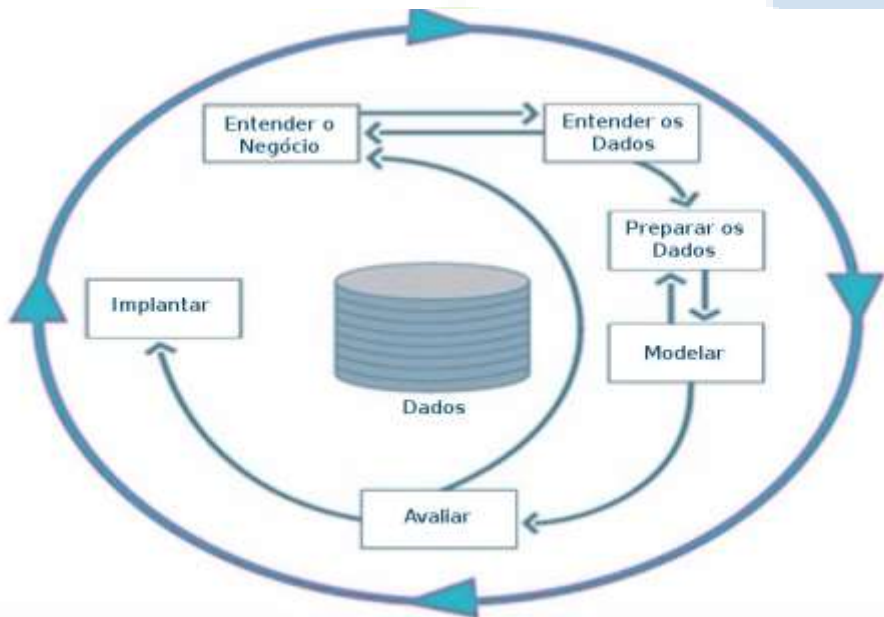






Ferramentas e Processos





CRISP-DM

Cross Industry Standard Process for Data Mining





Data Science Academy

Como Selecionar o Algoritmo Ideal Para Cada Problema?



Data Science Academy



- Árvores de decisão
- Random Forests
- Descoberta de associações e sequência
- Boosting e bagging de gradiente
- Máquinas de vetores de suporte
- Redes neurais
- Mapeamento de nearest-neighbor
- Cluster k-means
- Mapas auto-organizáveis
- Técnicas de otimização de busca local (por ex., algoritmos genéticos)
- Maximização da expectativa
- Análise Multivariada - Adaptive regression splines
- Redes Bayesianas
- Kernel para estimativa de densidade
- Análise de componentes principais
- Decomposição do valor singular
- Modelos de Gauss

São muitos os algoritmos
de Machine Learning





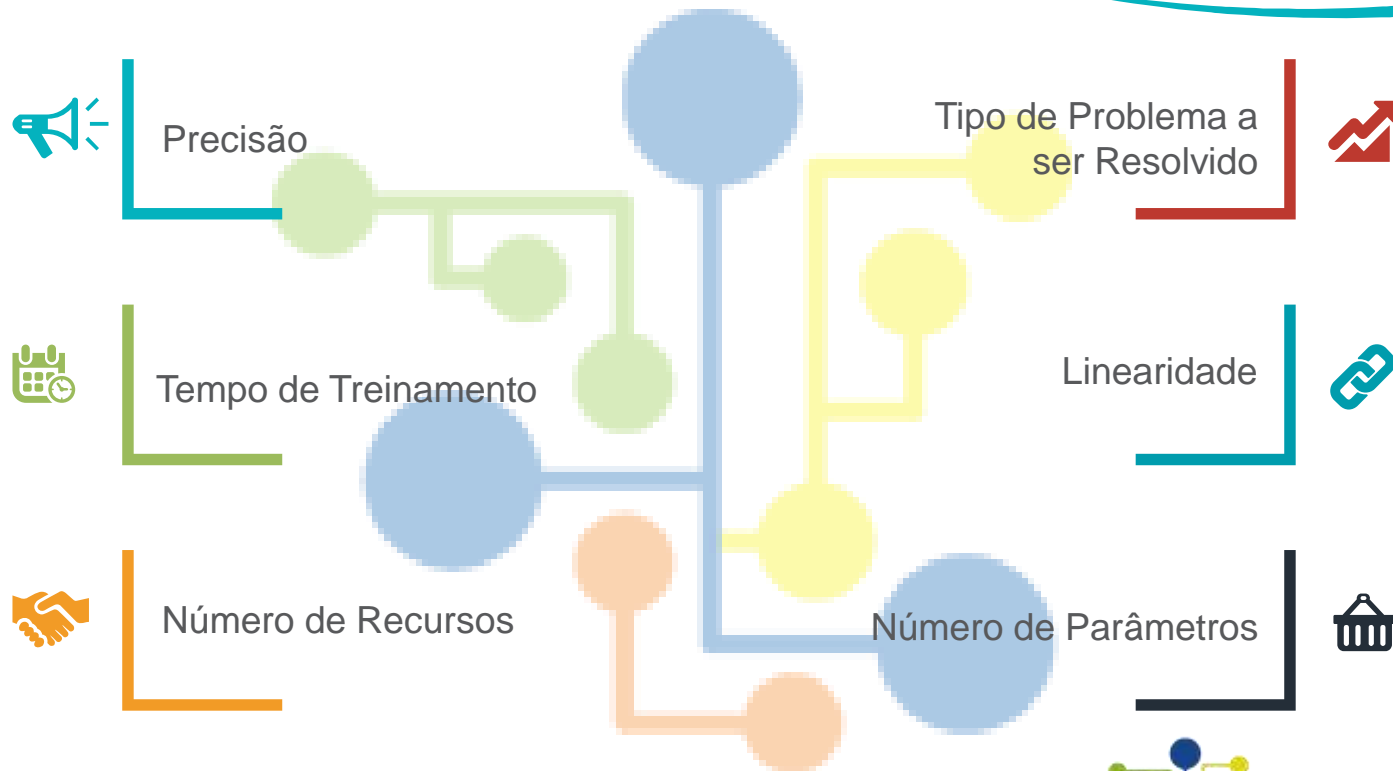
Quando alguém perguntar a você:

Qual algoritmo de Machine Learning devo usar?

A resposta correta será:

Depende.







Precisão





Tempo de
Treinamento





Linearidade





Número de
Parâmetros





Número de Recursos





Comparação entre os principais algoritmos





Classificação Binária (2 classes)

■ Alto
■ Moderado

Algoritmo	Tempo de Treinamento	Precisão	Linearidade
Regressão Logística	■	■	■
Árvore de Decisão	■	■	N/A
Random Forest	■	■	N/A
Redes Neurais	■	■	N/A
SVM	■	■	■
Métodos Bayesianos	■	■	■



Classificação Multiclasse (mais de 2 classes)

■ Alto
■ Moderado

Algoritmo	Tempo de Treinamento	Precisão	Linearidade
Regressão Logística	■	■	■
Árvore de Decisão	■	■	N/A
Random Forest	■	■	N/A
Redes Neurais	■	■	N/A
SVM	■	■	■





Regressão

■ Alto
■ Moderado

Algoritmo	Tempo de Treinamento	Precisão	Linearidade
Linear	■	■	■
Árvore de Decisão	■	■	N/A
Random Forest	■	■	N/A
Redes Neurais	■	■	N/A
Poisson	■	■	■





Não Supervisionados

■ Alto
■ Moderado

Algoritmo	Tempo de Treinamento	Precisão	Linearidade
K-Means	■	■	■
PCA	■	■	■





Data Science Academy

Obrigado