



Data Science
Academy

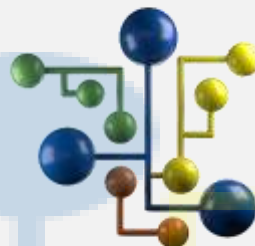
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

Machine Learning



Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

Seja muito bem-vindo(a)!





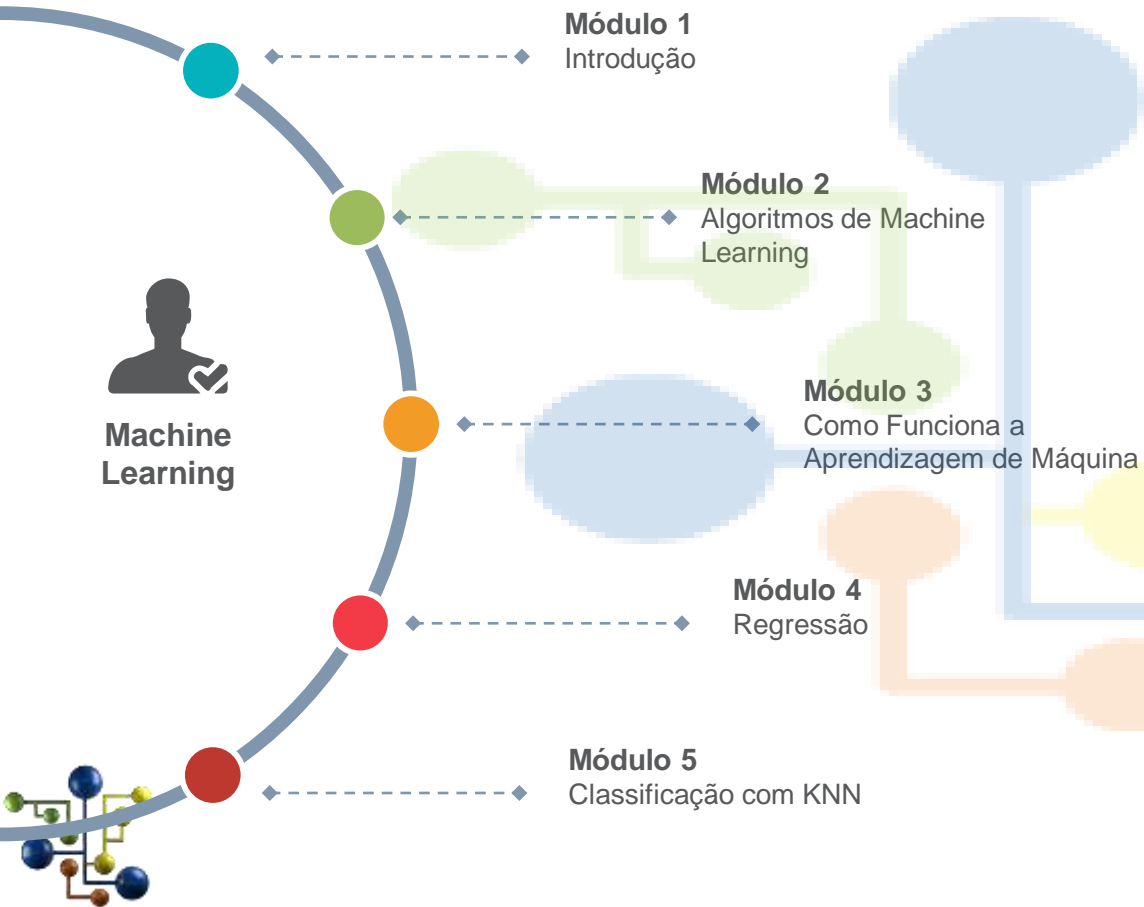
“

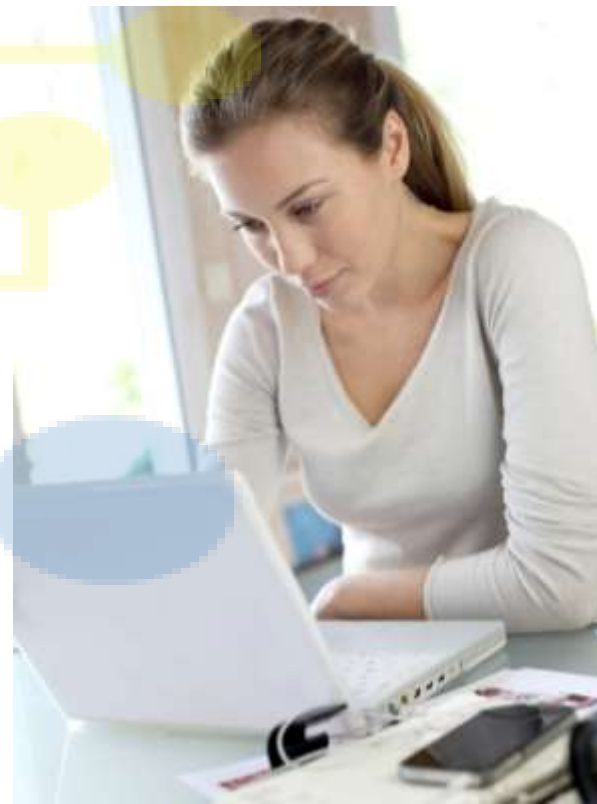
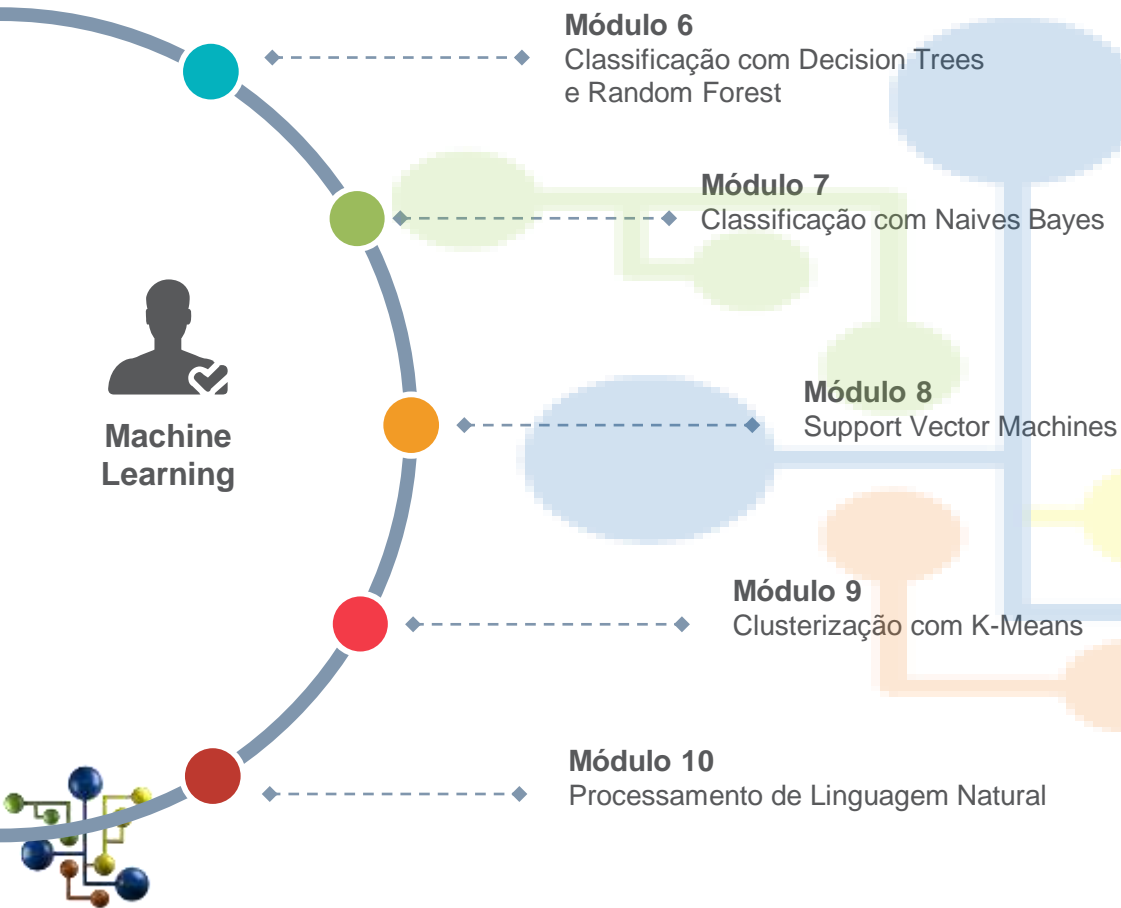
Machine Learning

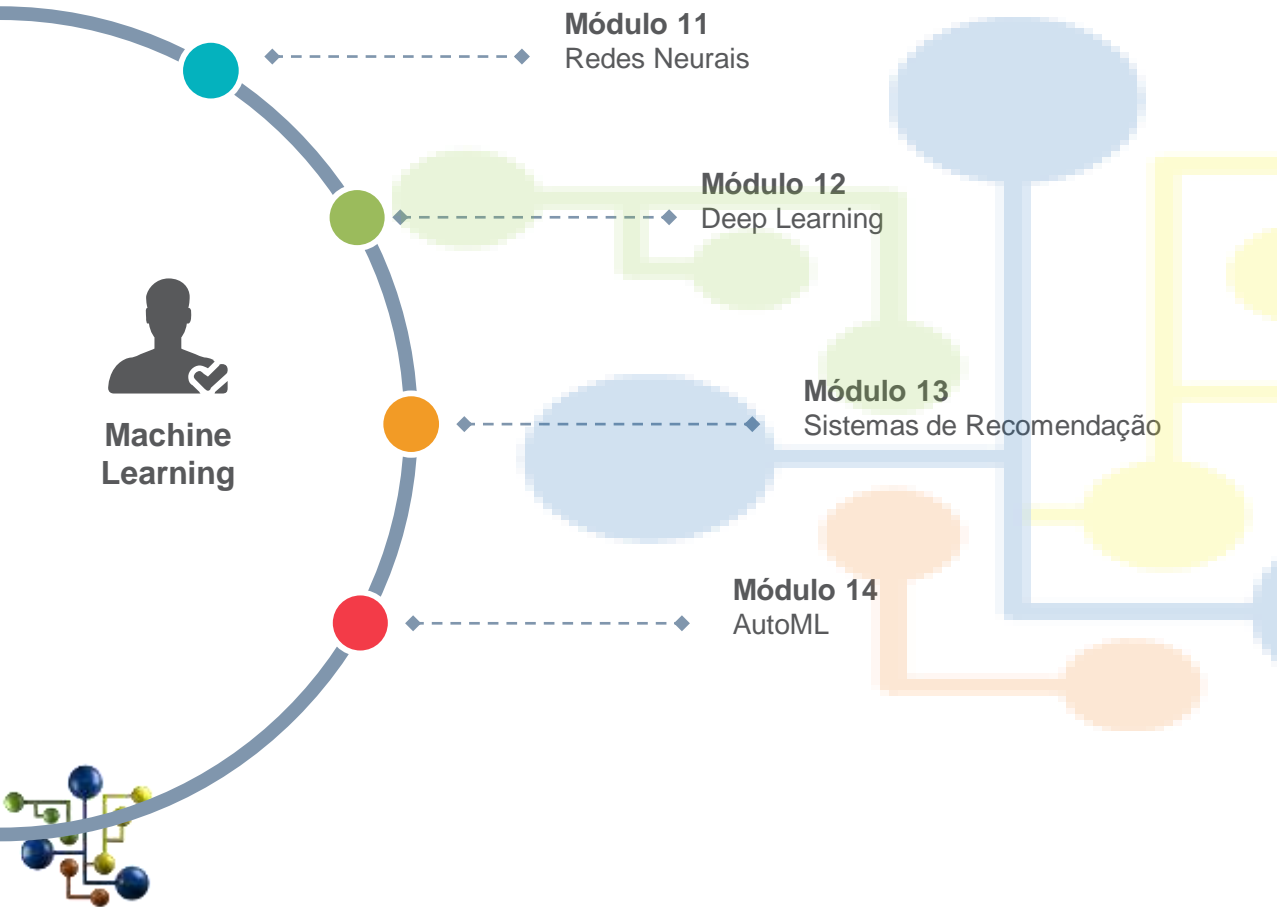
Introdução ao Curso

”





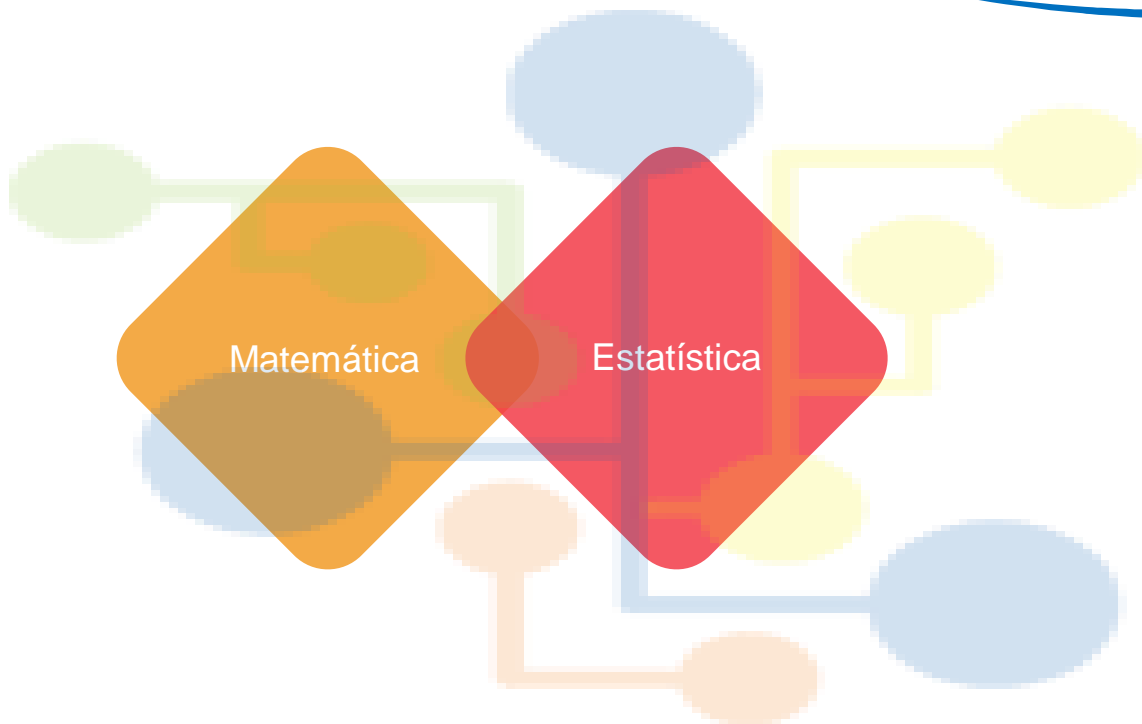






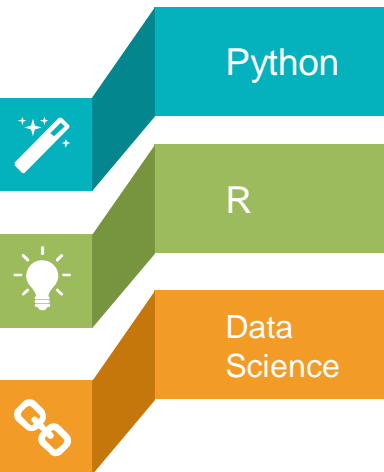
O que não veremos neste curso?







Pré-requisitos (recomendados)



- 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





Temos ainda a Formação *Análise Estatística* Para Cientistas de Dados com cursos de Matemática e Estatística!





Formação Cientista de Dados

Ou

Formação Análise Estatística Para Cientistas de Dados

Qual fazer primeiro?





Abordagem 1 – Você observa atentamente a figura que você quer montar (ou seja, define o objetivo na sua mente) que aqui chamaremos de “Big Picture”, e então começa a ver como as peças se encaixam. Com esta estratégia e a figura na sua mente, as peças individuais quando montadas vão fazendo cada vez mais sentido e mesmo quando você olha para uma peça isolada, seu cérebro tenta associá-la com a Big Picture.



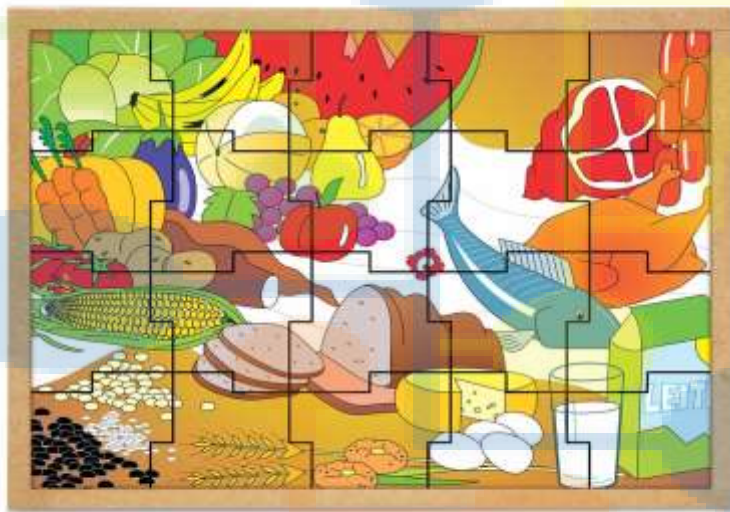


Abordagem 2 – Você não observa a figura (ou muitas vezes nem mesmo tem a figura pronta) e então vai investigando as peças individualmente e tentando descobrir as ligações, até que alguma figura faça sentido.





Qual das duas abordagens é a mais rápida?

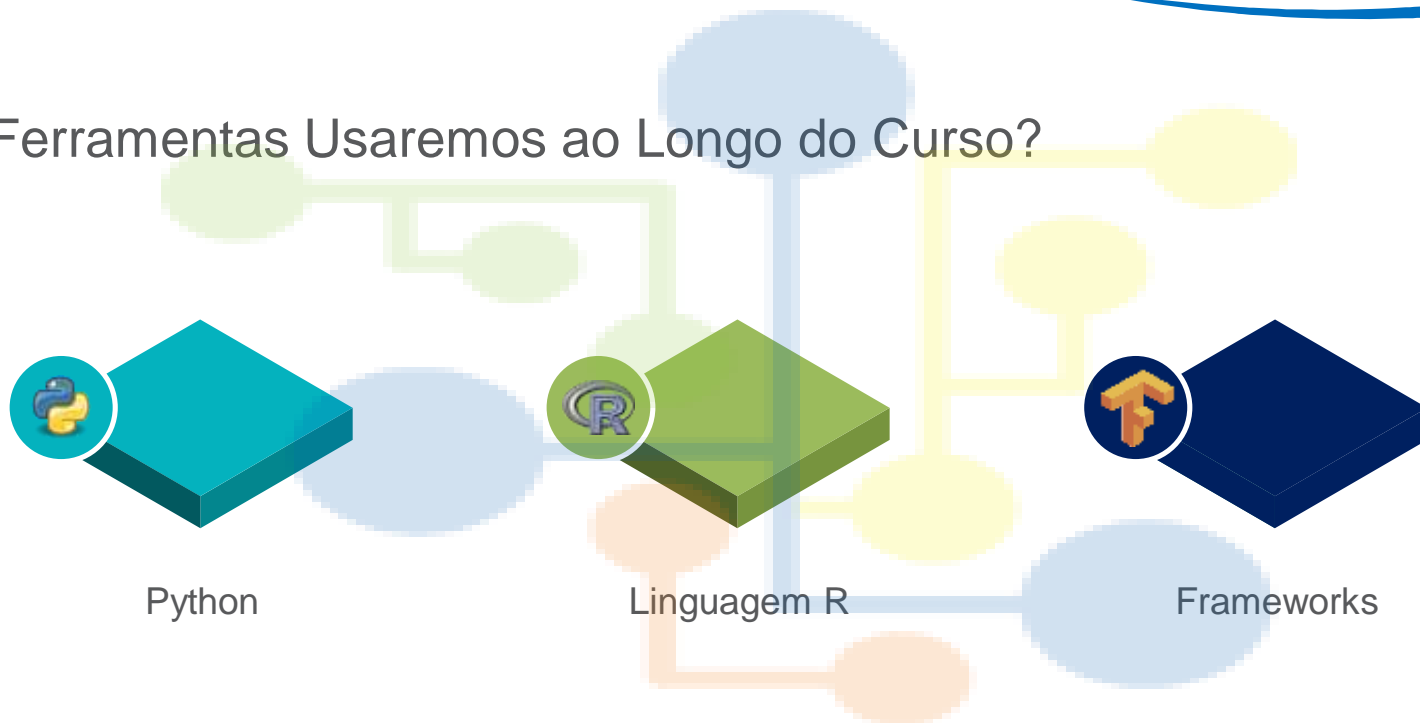


A abordagem 1 é mais rápida e mais eficiente!





Quais Ferramentas Usaremos ao Longo do Curso?





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!



Objetivos ao fim deste curso



100%

Desenvolver o processo
de modelagem de dados
para Machine Learning





Objetivos ao fim deste curso



100%

Conhecer os 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





Avaliação Final



50 Questões



3 Tentativas

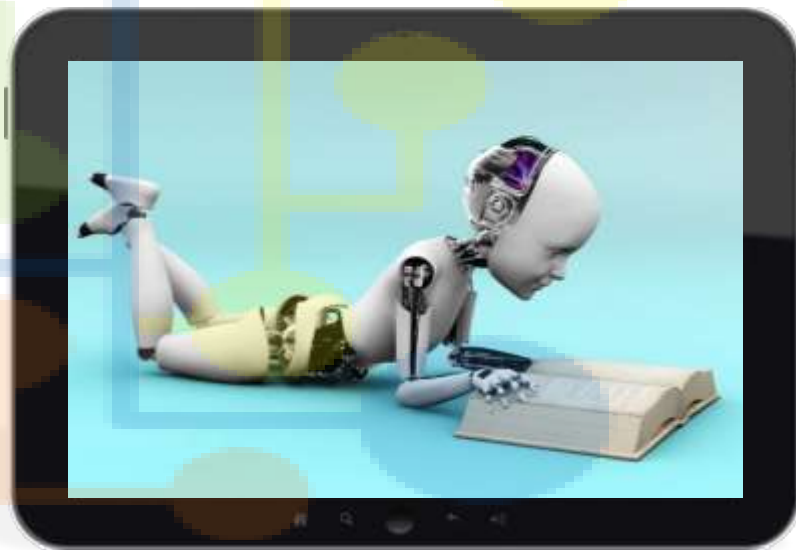


70% de Aproveitamento





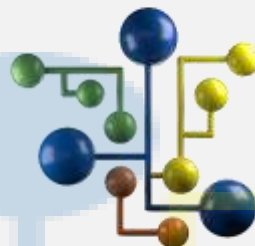
Acesse o Curso do Smartphone ou Tablet com
nossas Apps para iOS e Android





Data Science
Academy

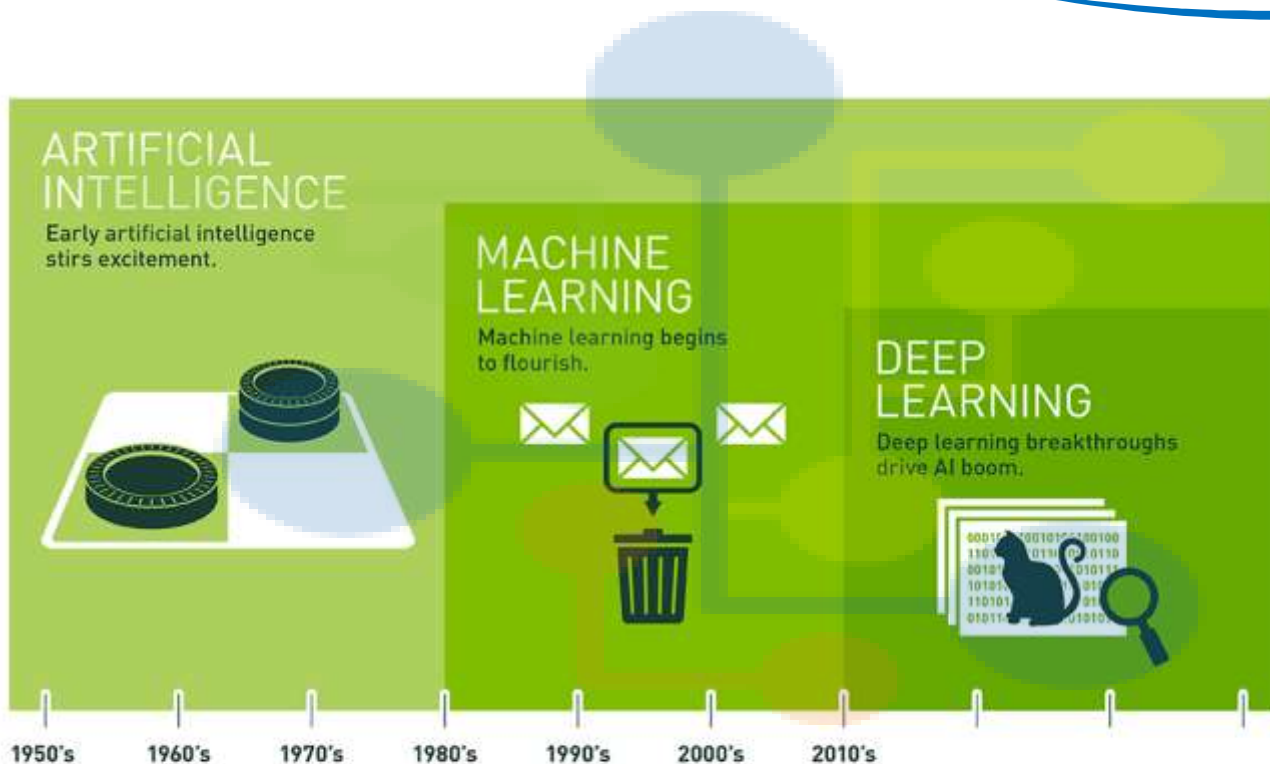
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

O que é Aprendizado de Máquina?





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.





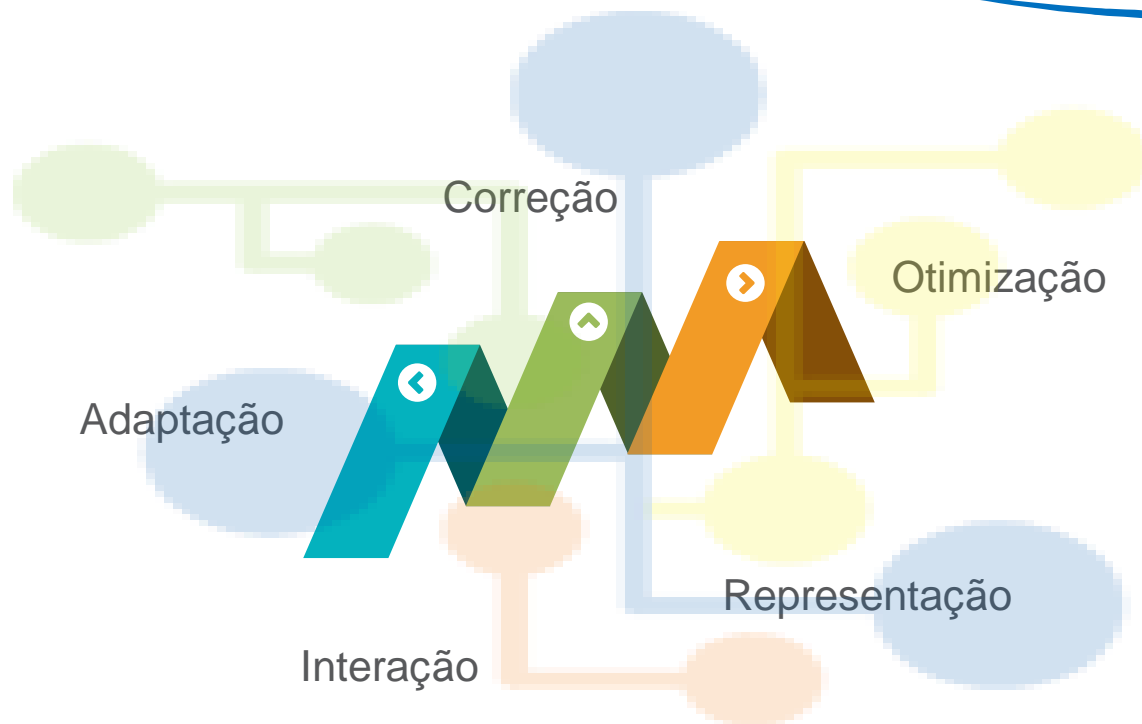
O que é Aprendizado?





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.







Há grande semelhança entre o 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!

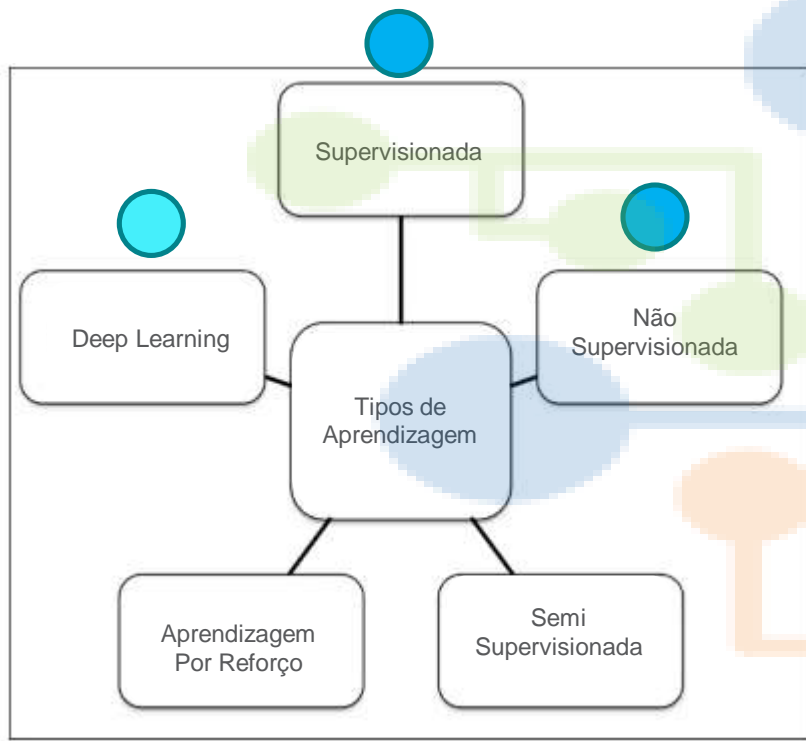






Machine Learning ou Aprendizado de Máquina é um método de análise de dados que automatiza o desenvolvimento de modelos analíticos. Usando algoritmos que aprendem iterativamente a partir de dados, o aprendizado de máquina permite que os computadores encontrem insights ocultos sem serem explicitamente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



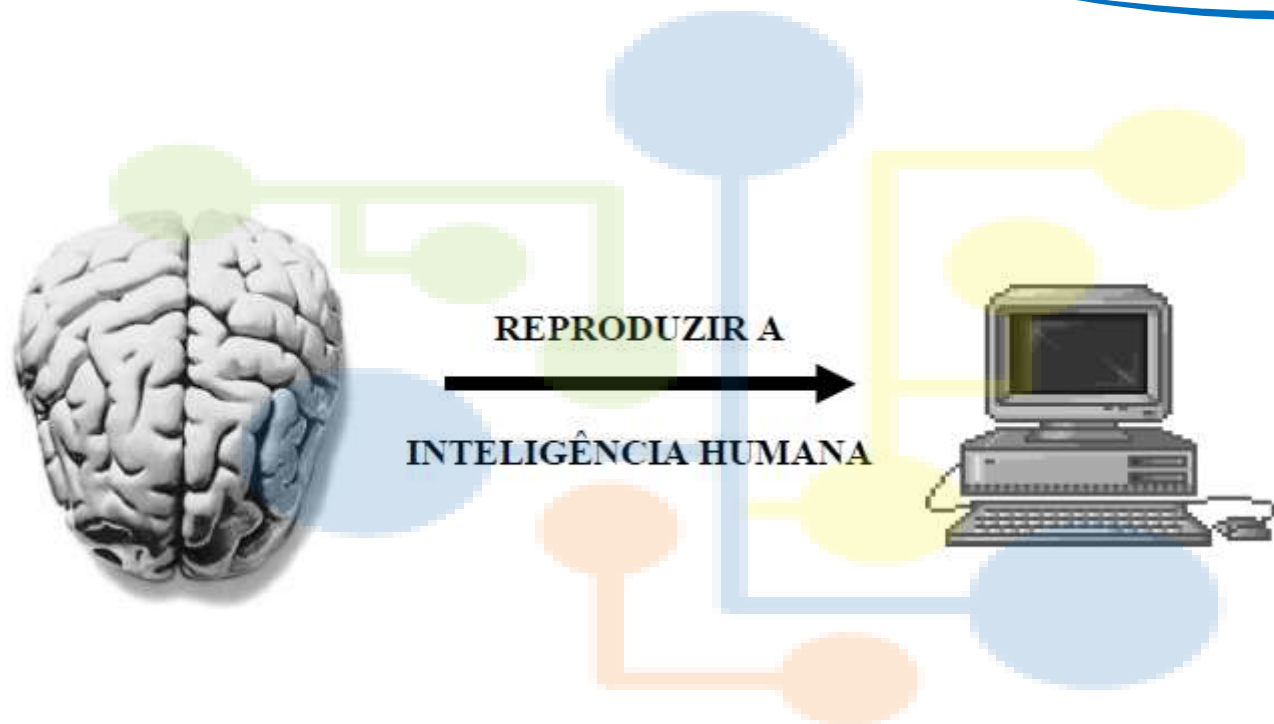


Tipos de Aprendizagem



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







Inteligência

A decorative background diagram featuring a central red rounded rectangle with the text 'Dotado de inteligência, capaz de compreender, esperto, habilidoso'. Surrounding this rectangle are several colored circles (blue, green, yellow, orange) connected by lines, forming a network-like structure.

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.

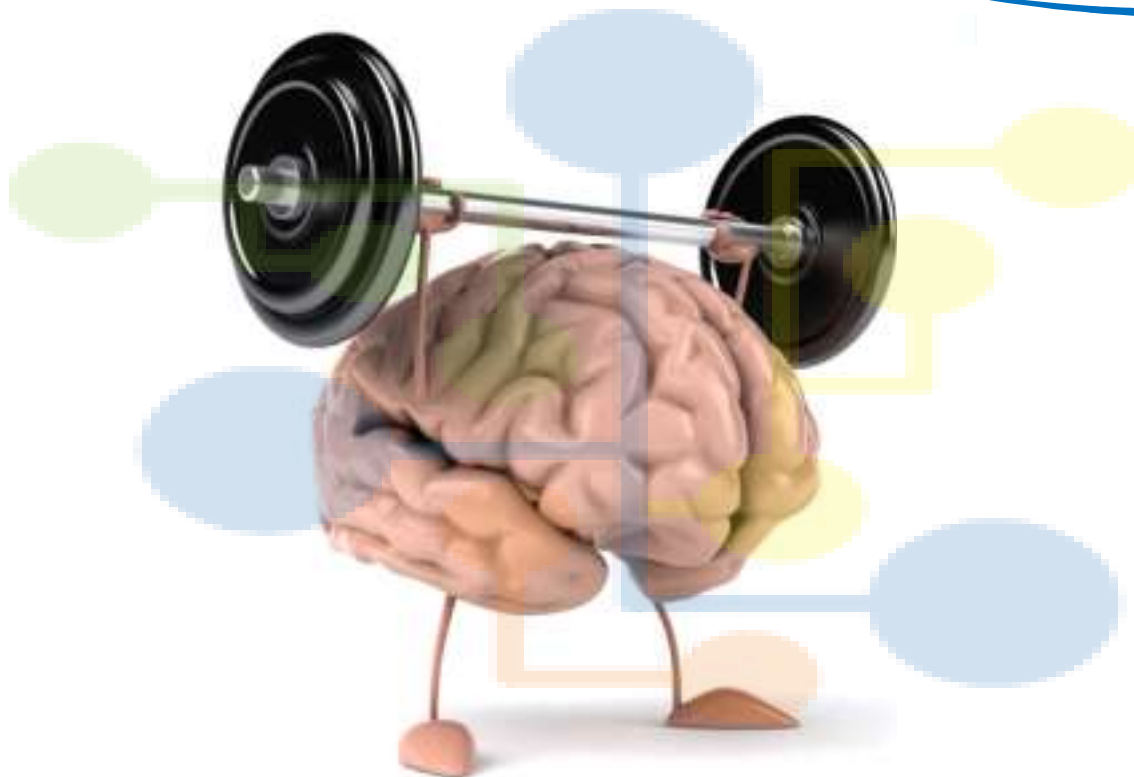




Inteligência Artificial

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!





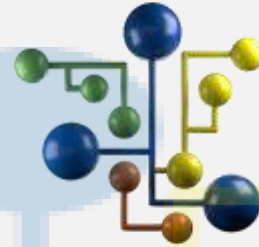
Don't model the World; Model the Mind.





Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

Por que Machine Learning Está Transformando o Mundo?

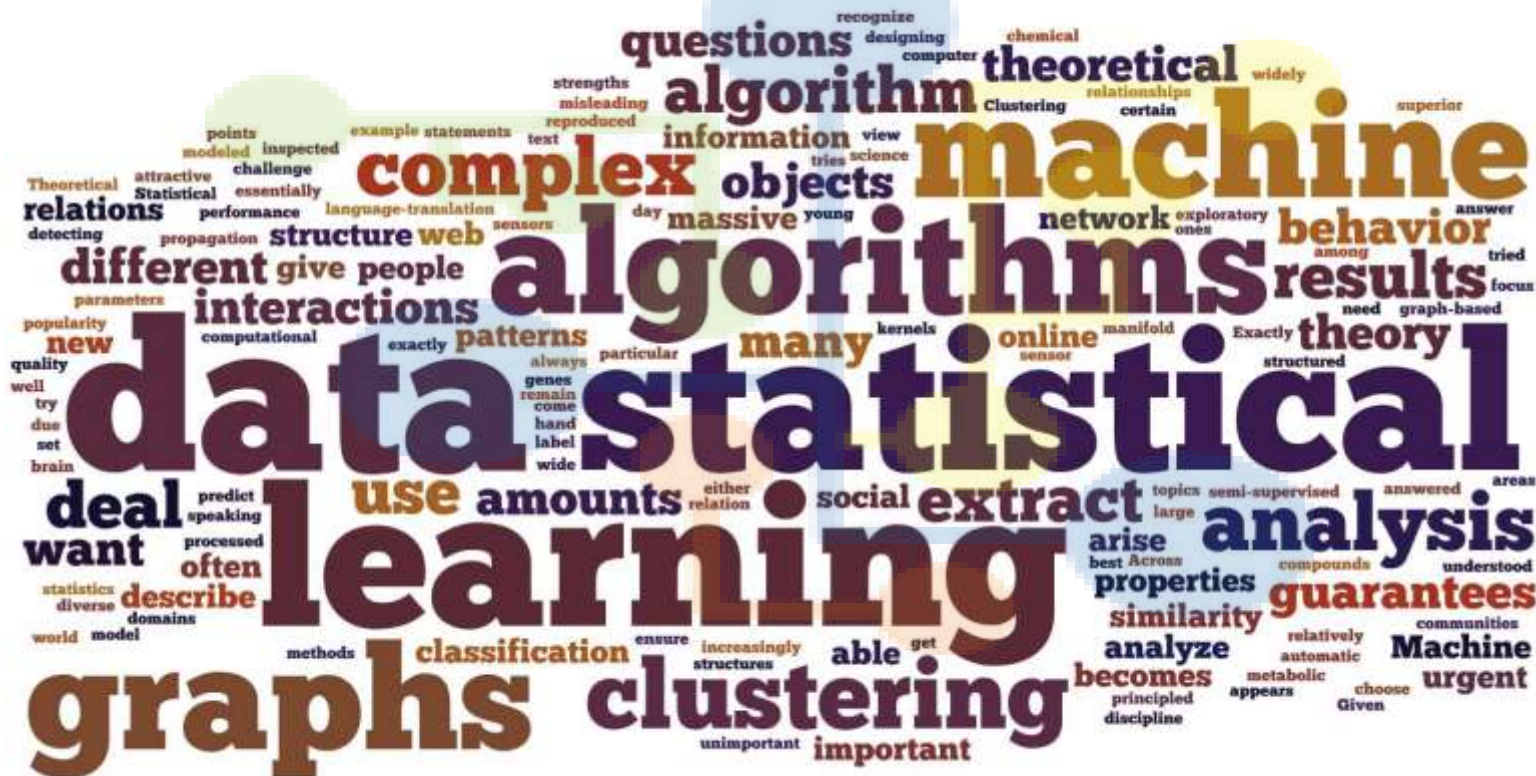


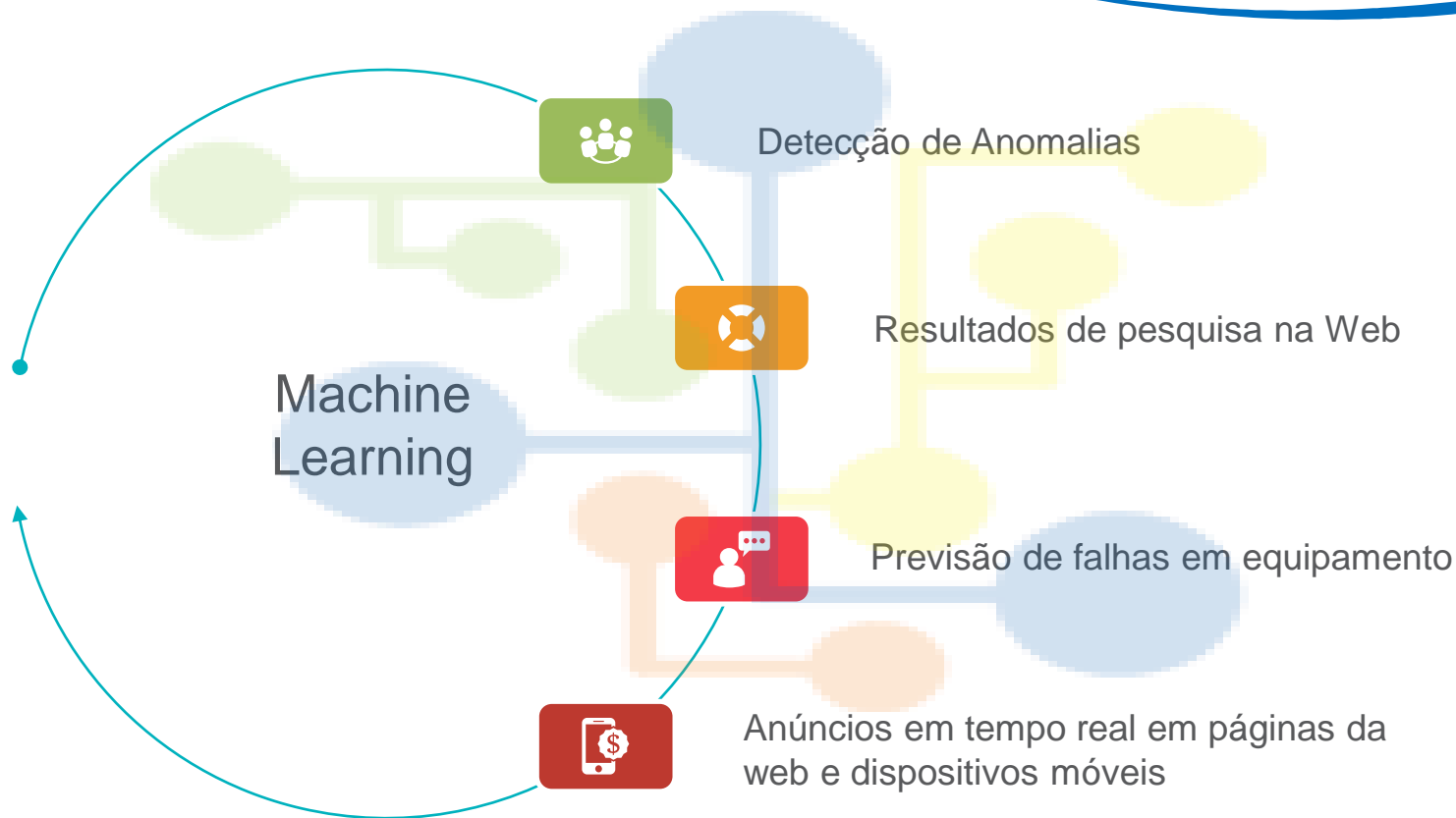


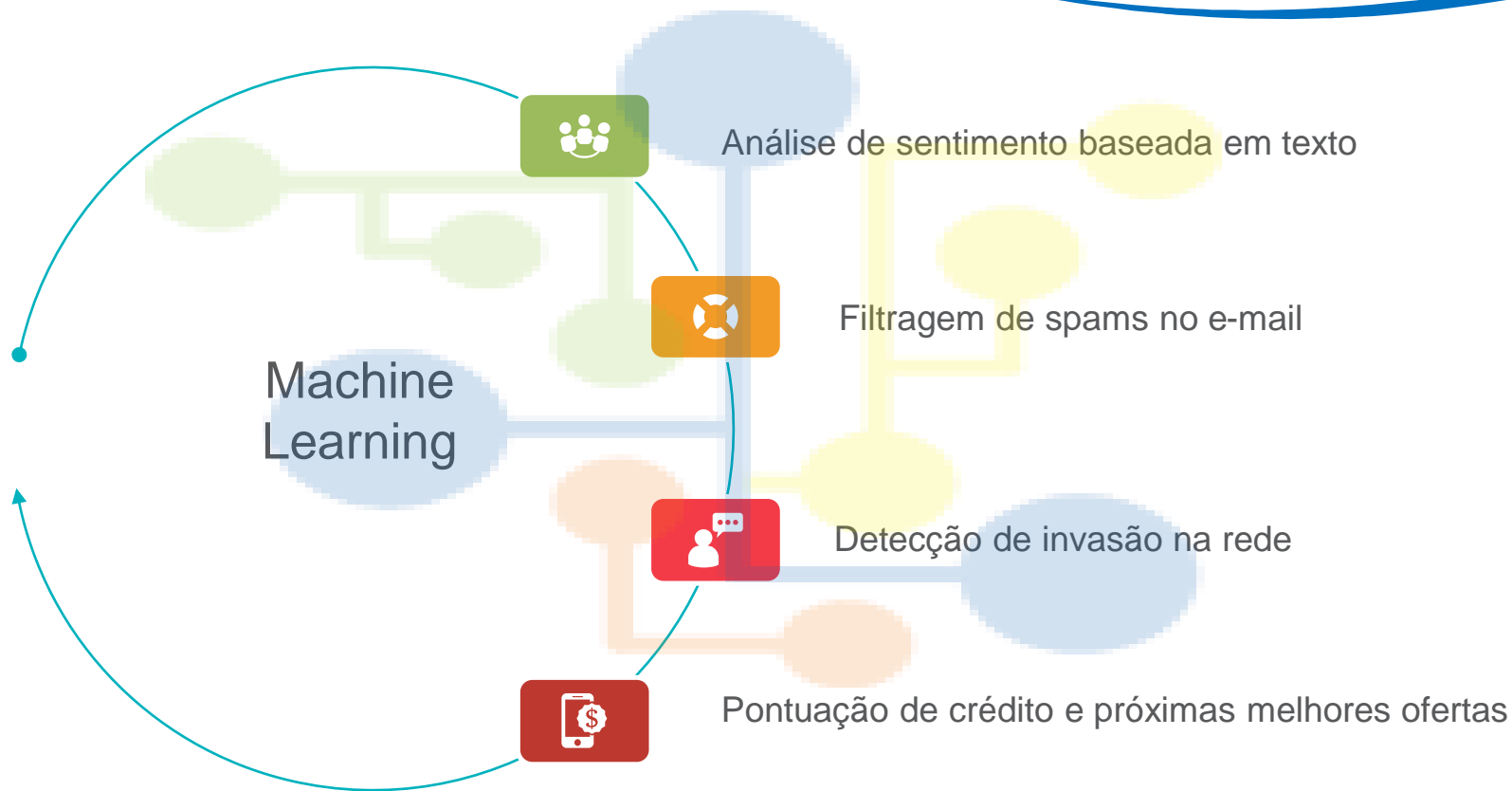
Algoritmos de aprendizagem de máquina, aprendem a induzir uma função ou hipótese capaz de resolver um problema a partir de dados que representam instâncias do problema a ser resolvido.

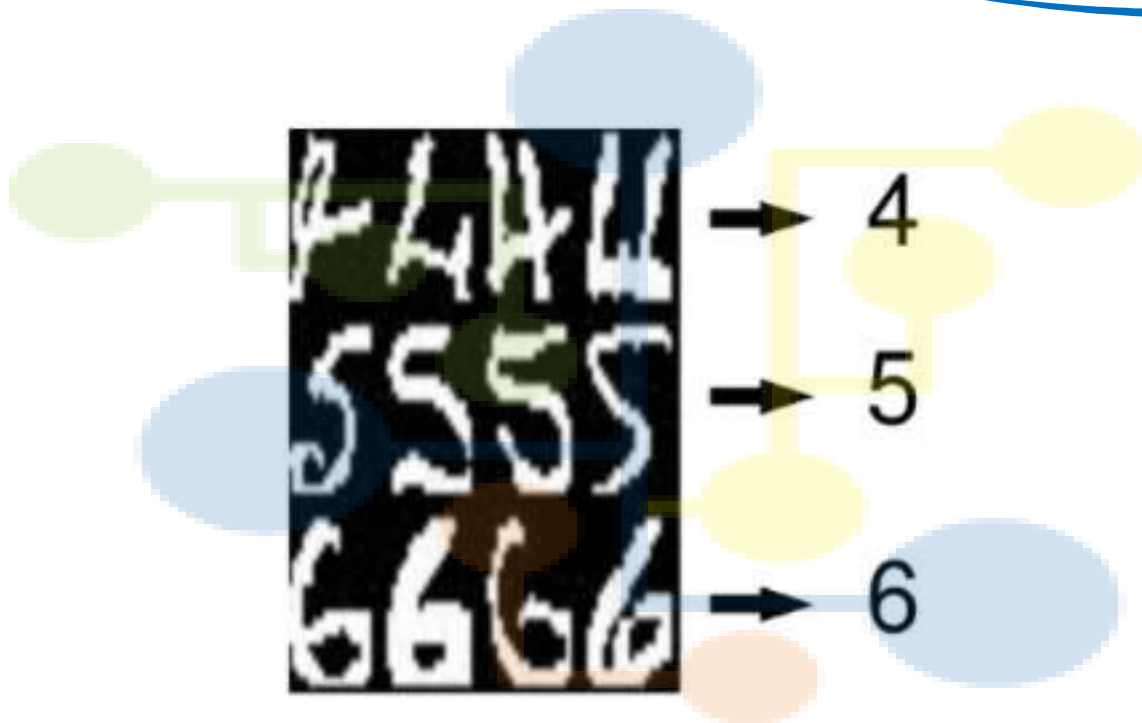














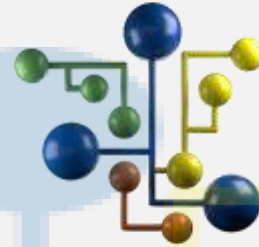
Machine Learning não está transformando nosso mundo;
Machine Learning já transformou o nosso mundo.





Data Science
Academy

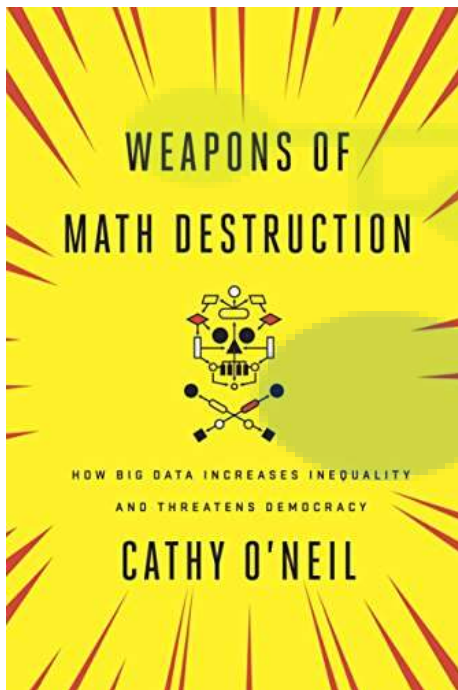
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

The Dark Side of Big Data





Recomendo

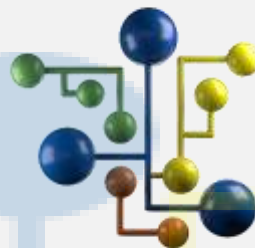
Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy





Data Science
Academy

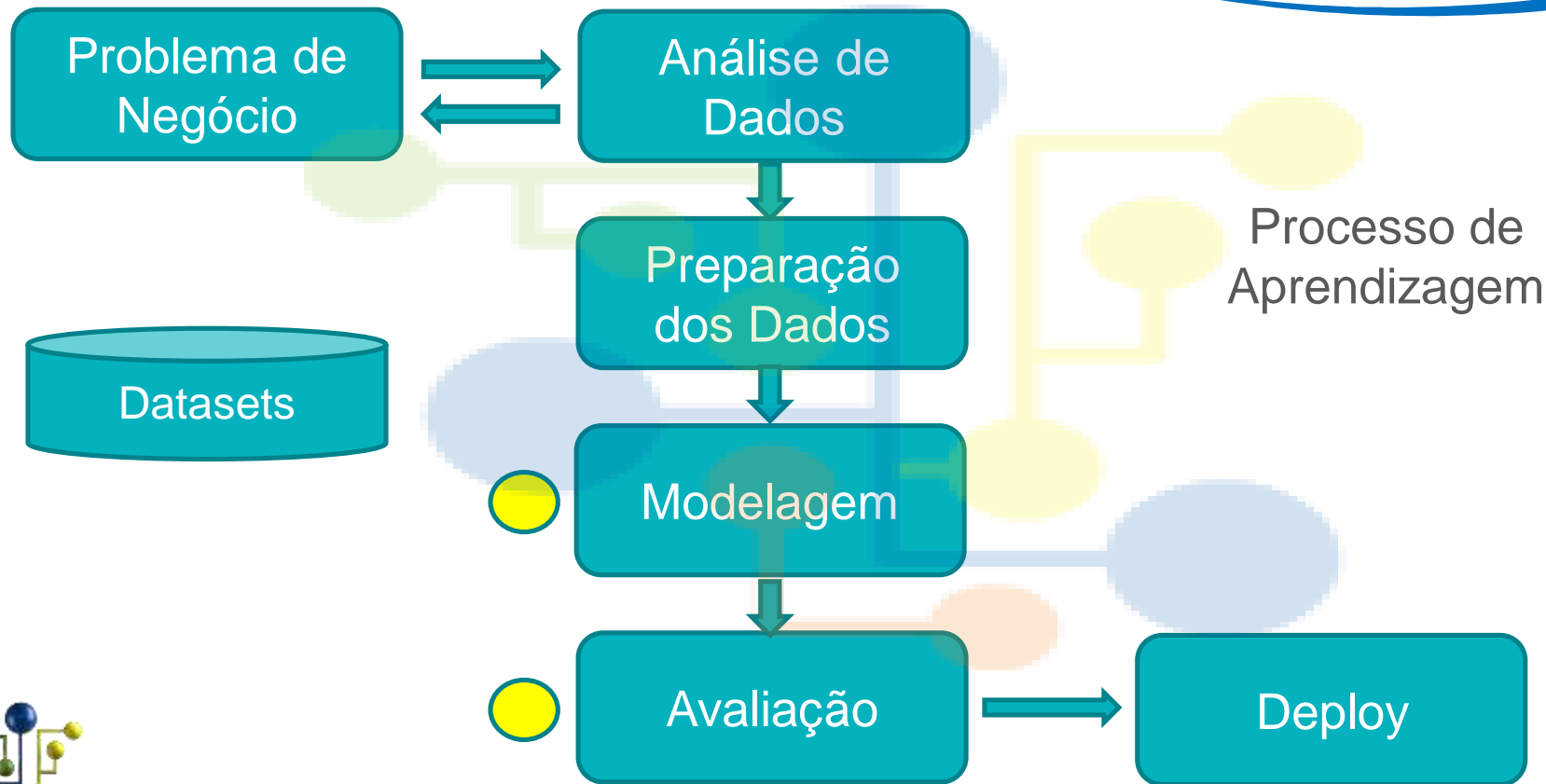
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

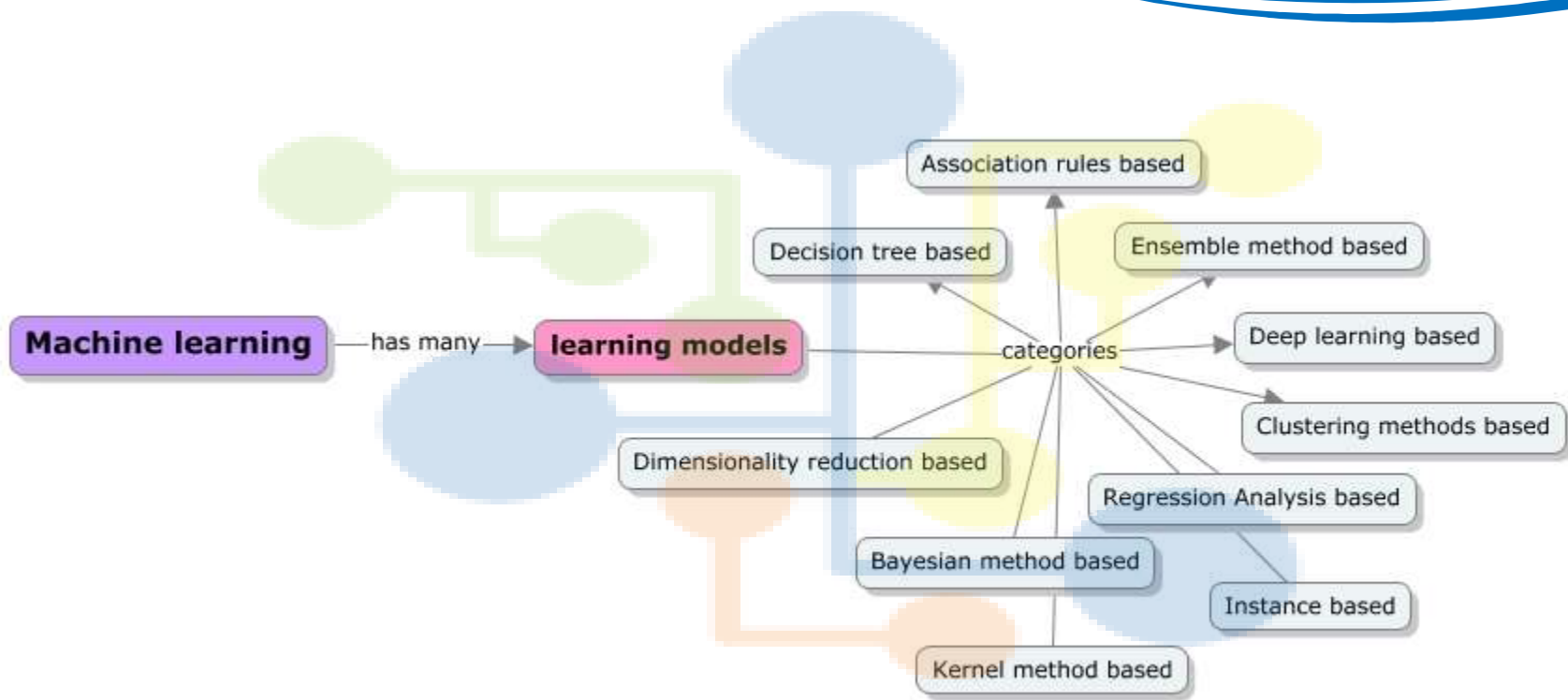


**Data Science
Academy**

Processo de Aprendizagem



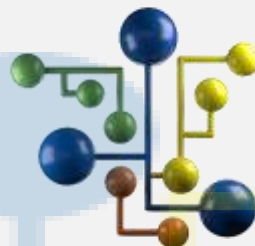






Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

Tipos de Aprendizagem





A background diagram showing a central blue circle connected to three other circles (green, yellow, and orange) by lines. The green circle is connected to the blue circle by a green line, the yellow circle by a yellow line, and the orange circle by an orange line. The lines are thick and have a slight 3D effect.

Aprendizagem
Supervisionada

Aprendizagem
Não
Supervisionada

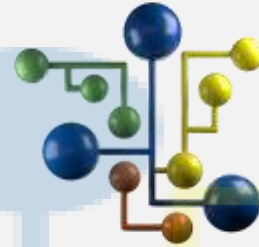
Aprendizagem
Por Reforço





Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

Aprendizagem Supervisionada





Aprendizagem Supervisionada





Aprendizagem Supervisionada

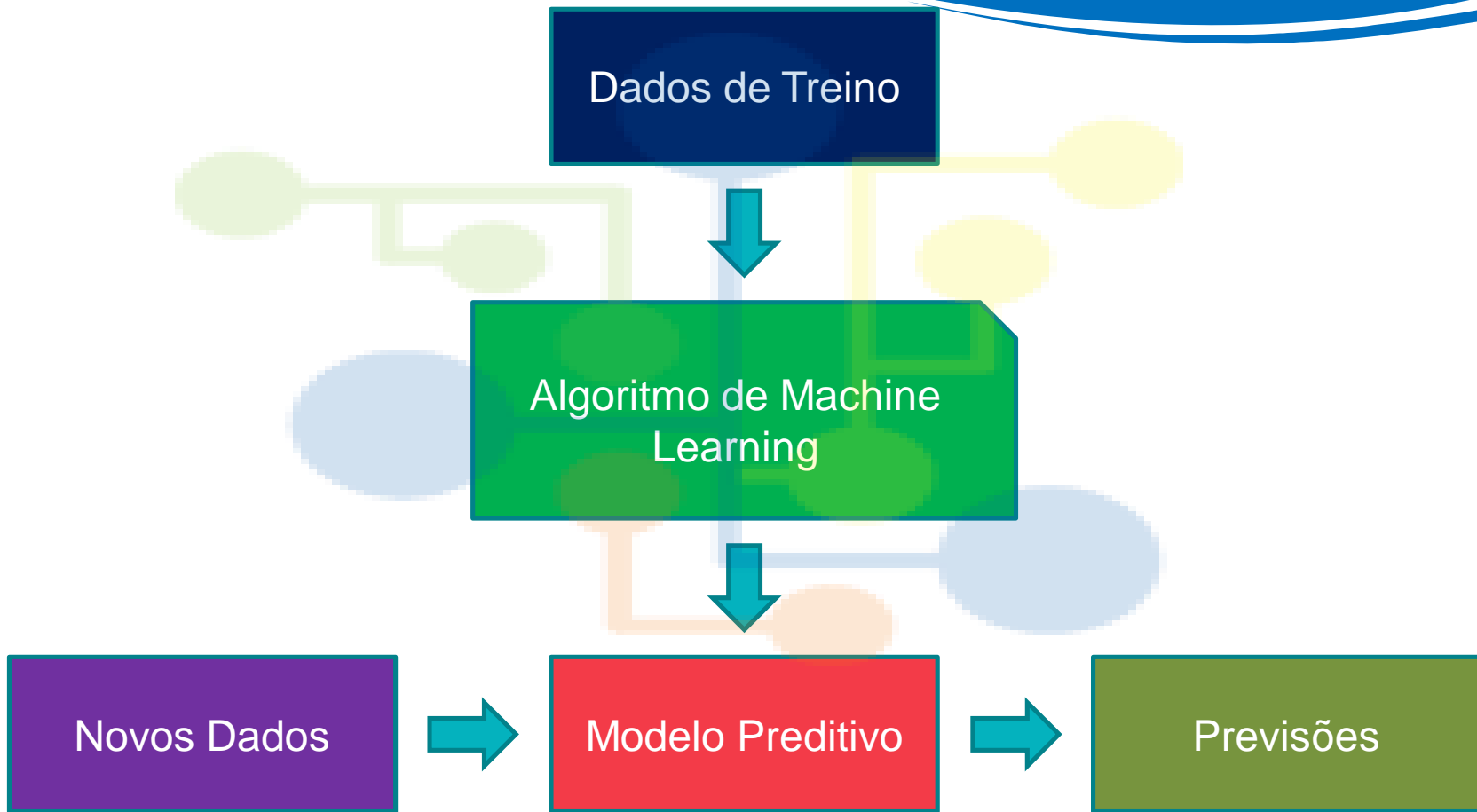
É o termo usado sempre que o programa é “treinado” sobre um conjunto de dados pré-definido.





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







Aprendizagem Supervisionada

Classificação

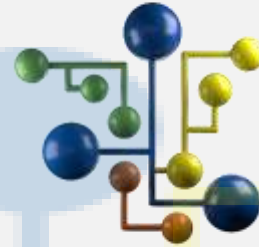
Regressão





Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

Aprendizagem Não Supervisionada





Aprendizagem Supervisionada





Aprendizagem Não Supervisionada





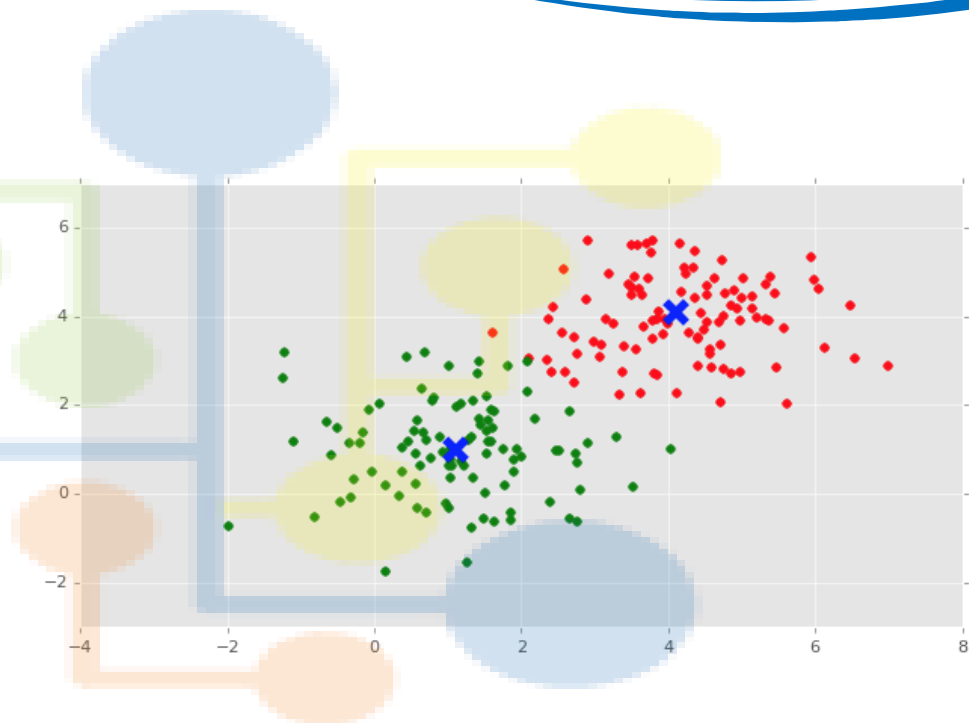
Aprendizagem Não-Supervisionada

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





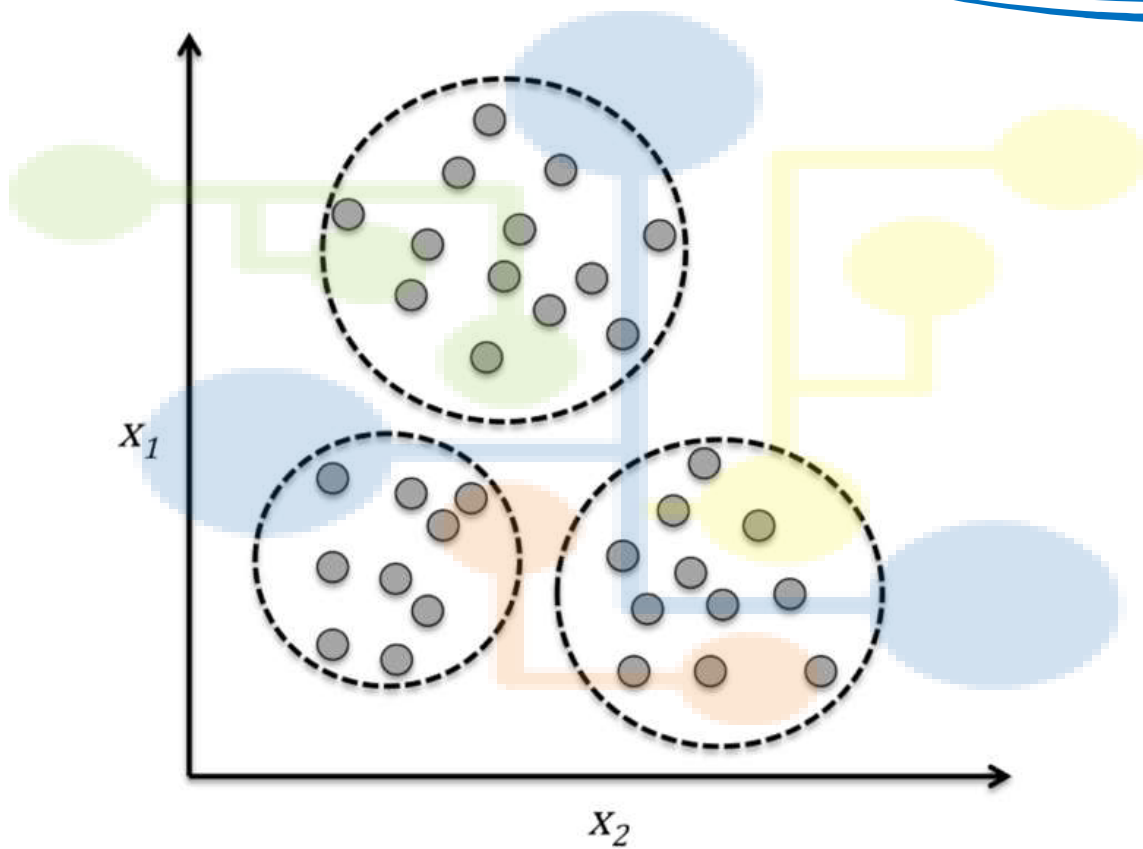
O objetivo de um algoritmo de aprendizado não supervisionado é organizar os dados de alguma forma ou descrever sua estrutura.





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







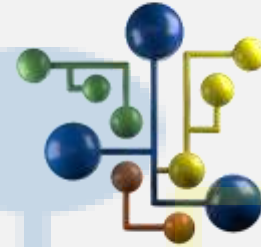
Aprendizagem Não-Supervisionada

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





Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

Aprendizagem Por Reforço (Reinforcement Learning)

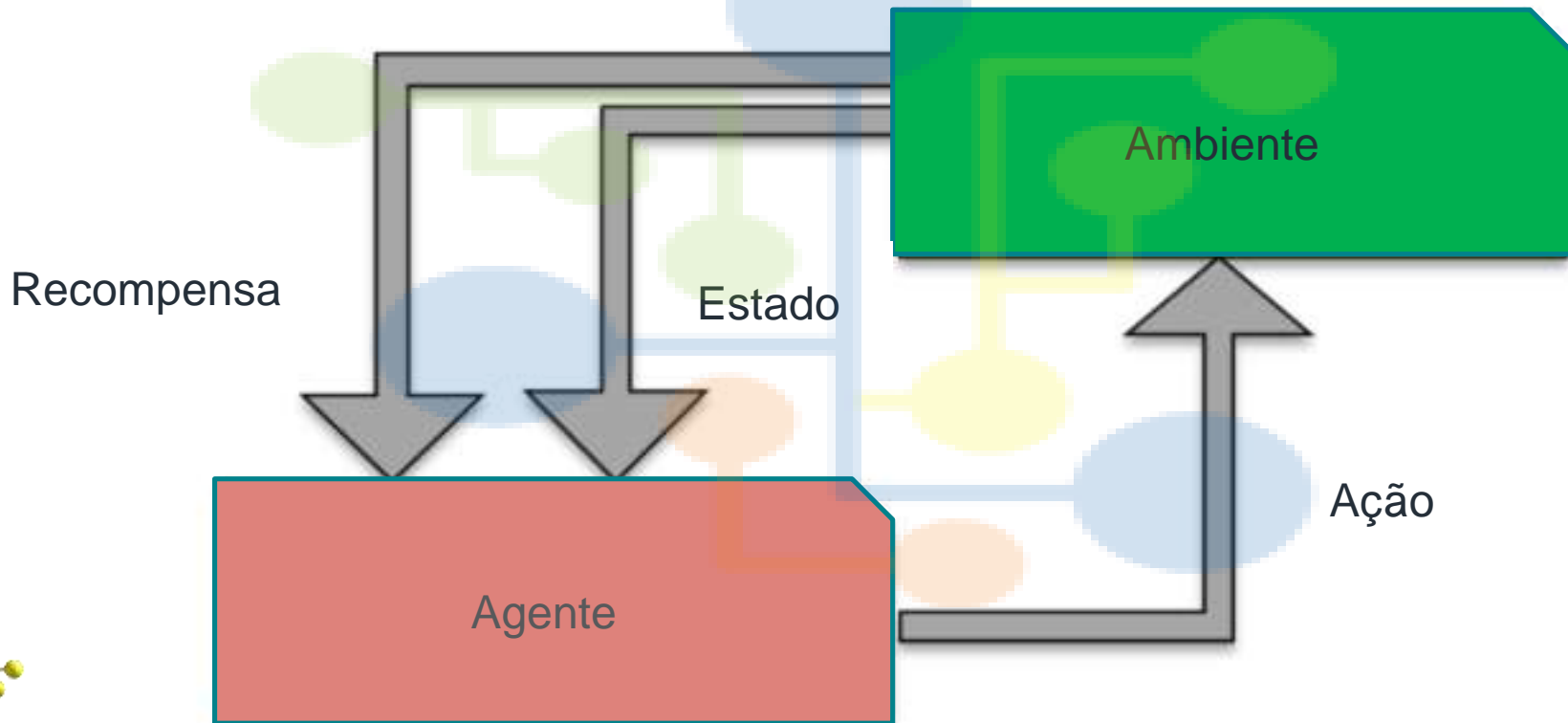




Reinforcement Learning

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.





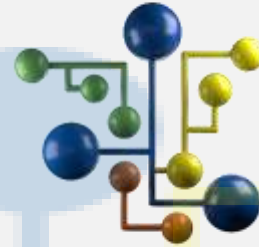
A ideia básica é simplesmente capturar os aspectos mais importantes do problema real que um agente de aprendizagem enfrenta durante a interação com o ambiente para alcançar uma meta.





Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

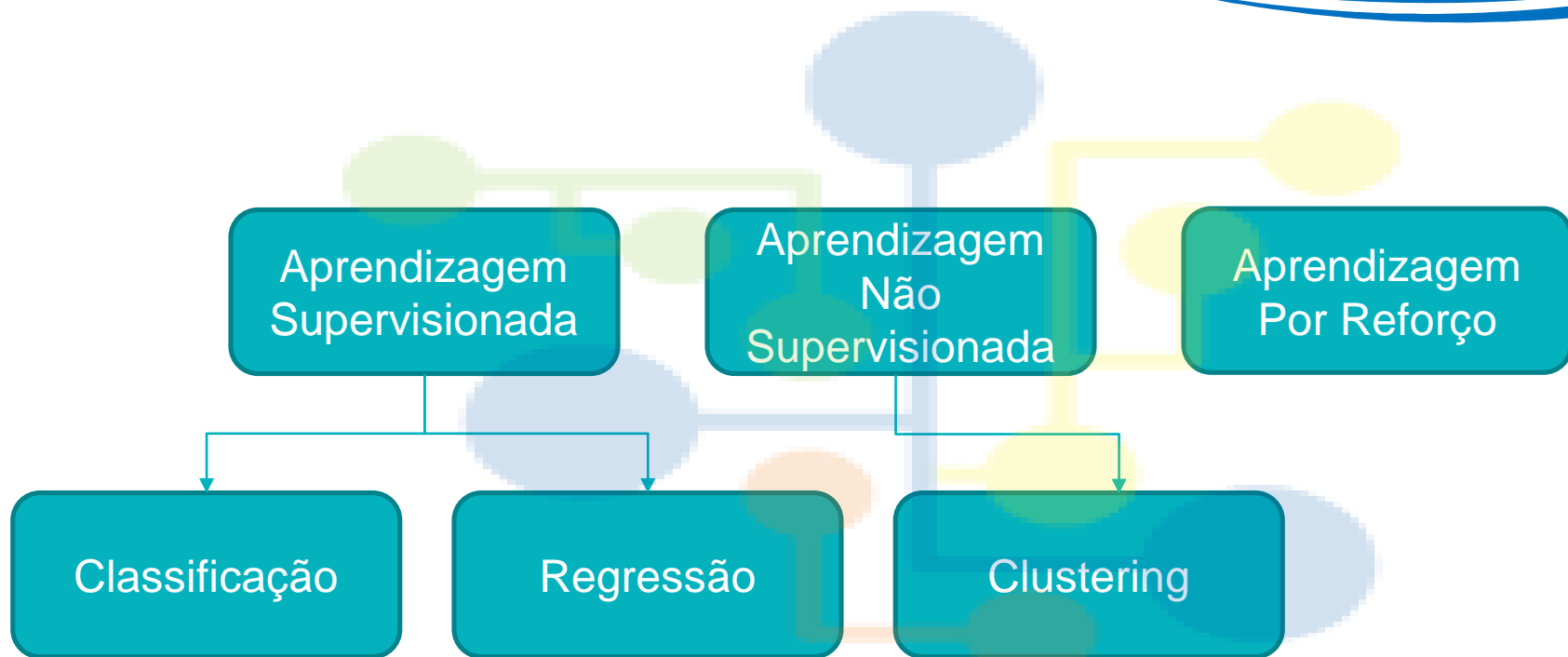


**Data Science
Academy**

Aprendizagem Supervisionada

Classificação







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

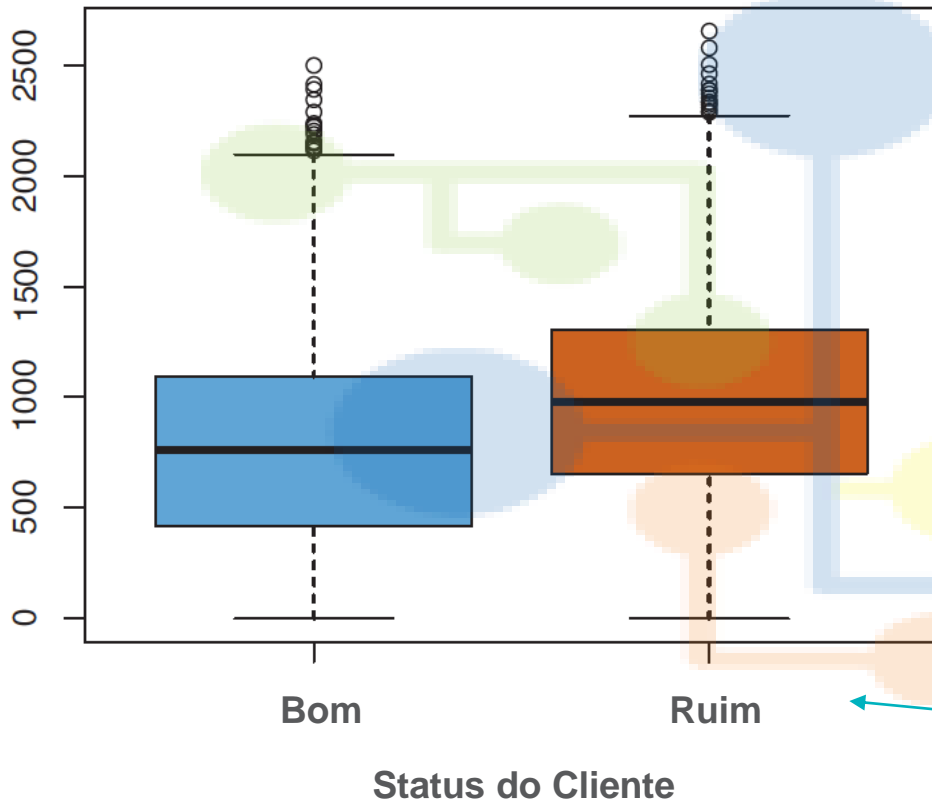
Exemplo: determinar o diagnóstico de uma doença em um paciente, observando as características similares em outros grupos de pacientes.





Variável
Preditora

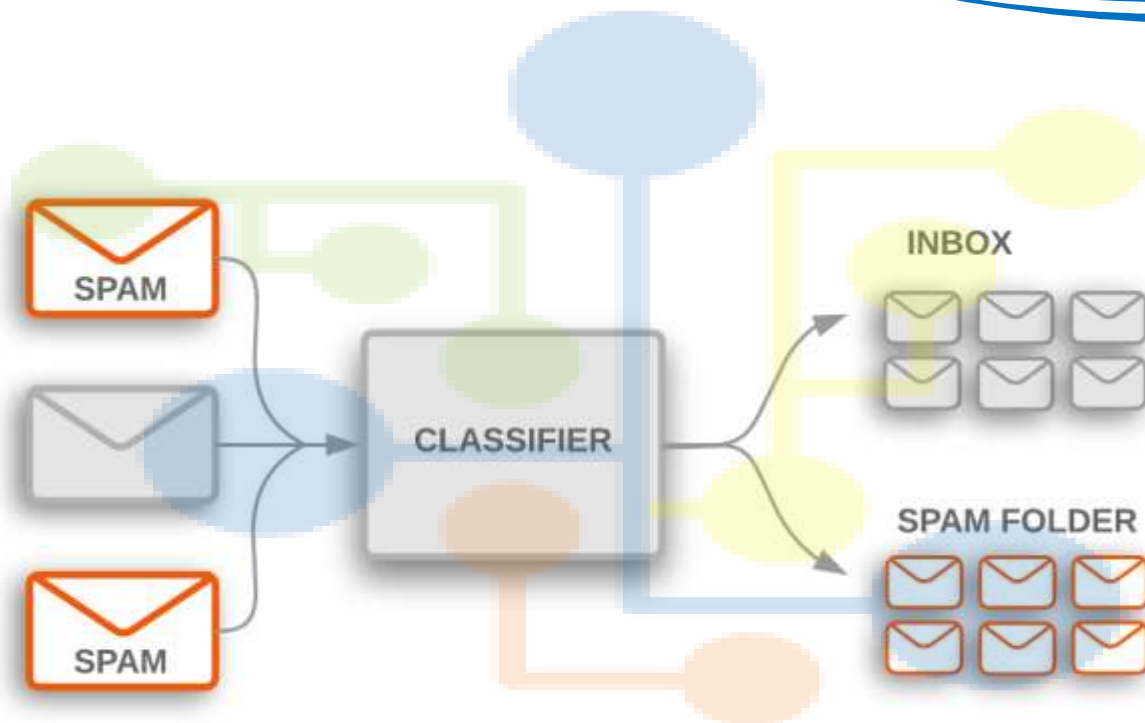
Saldo da Conta Corrente



Classificação

Variável Target
Pode assumir os valores:
Bom ou Ruim

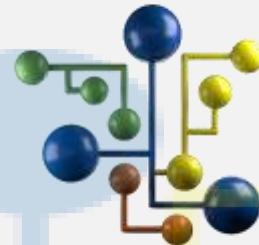






Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

Aprendizagem Supervisionada

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).

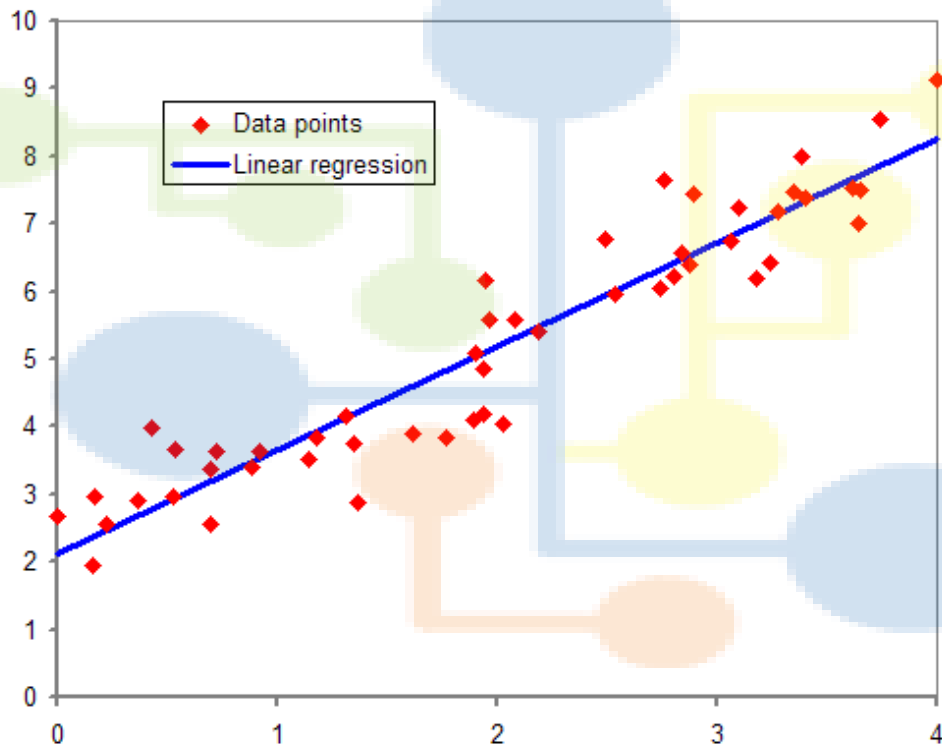


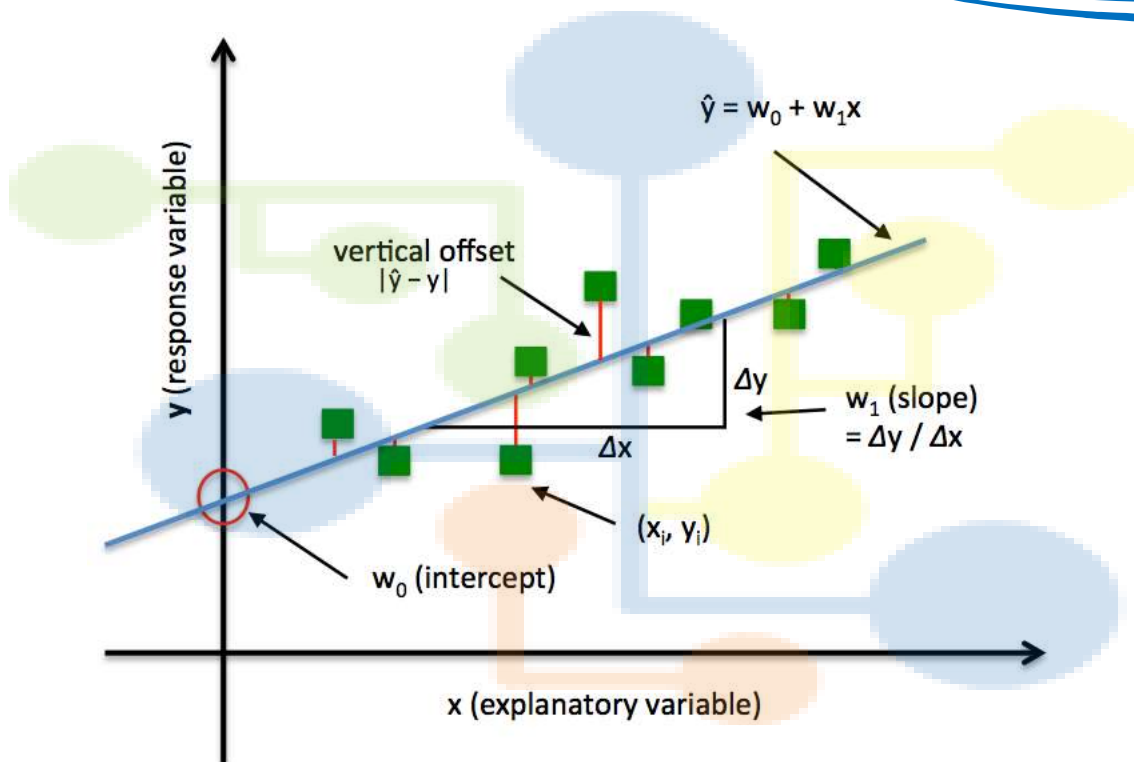


Como a Regressão pode ser usada?

- Investigação Científica
- Relações Causais
- Identificação de Padrões







A large, faint, stylized diagram in the background, resembling a circuit board or a network graph. It features several interconnected nodes of different colors (blue, green, yellow, orange) and lines connecting them, creating a complex, abstract shape.

Correlação Não Implica Causalidade!





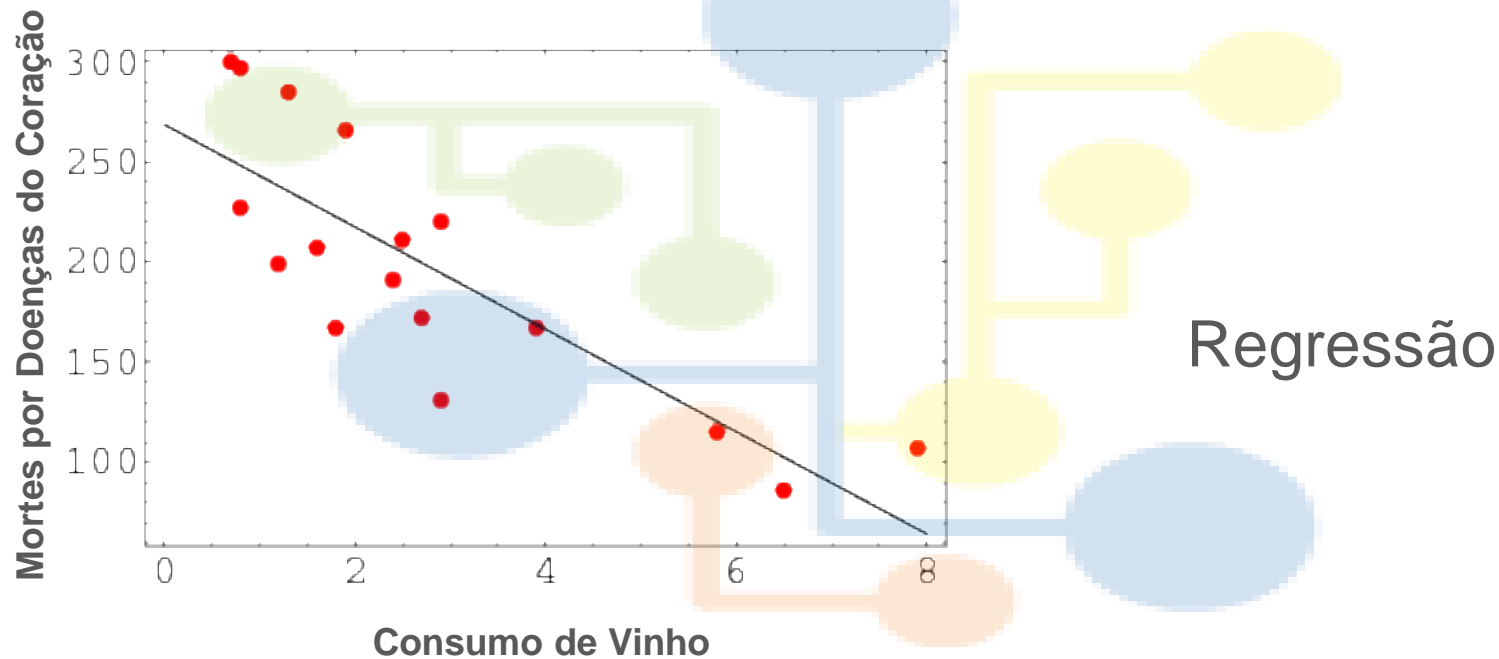
Só porque (A) acontece juntamente com (B)
não significa que (A) causa (B).





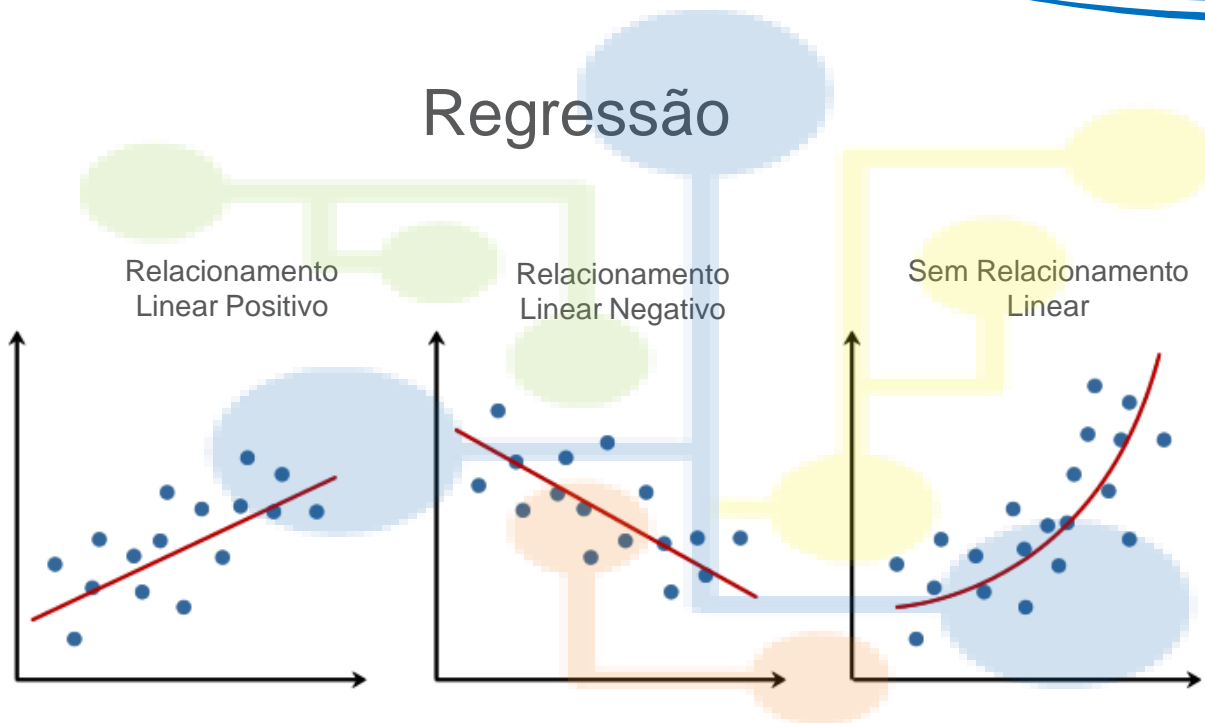
Regressão







Regressão





Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

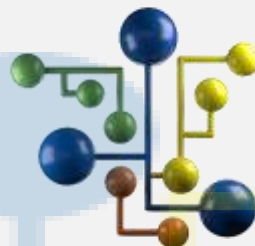
Machine Learning





Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

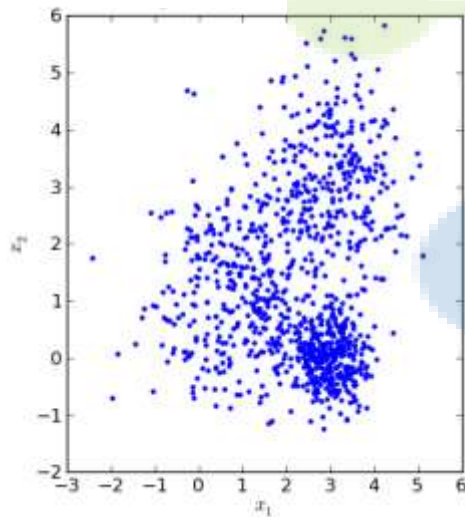
Aprendizagem Não-Supervisionada

Clusterização

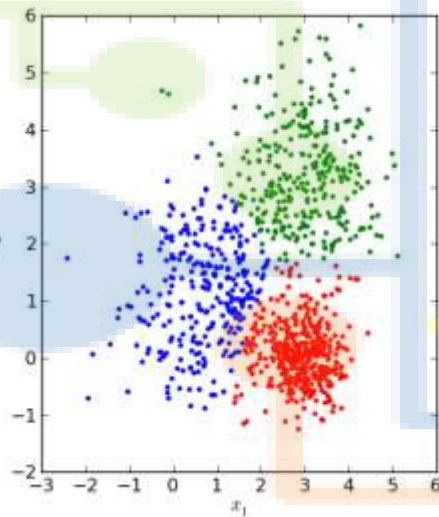




Antes da Clusterização



Depois da Clusterização



Clustering





Algoritmos de Aprendizagem Não Supervisionada

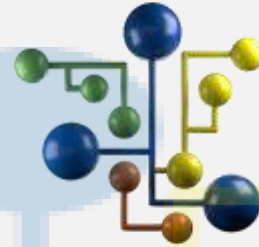
Categoria	Algoritmo
Algoritmos Baseados em Centroides	K-means, Gaussian Mixture Model, Fuzzy c-mean
Algoritmos Baseados em Conectividade	Algoritmos hierárquicos
Algoritmos Baseados em Densidade	DBSCAN, Optics
Probabilísticos	LDA
Redução de Dimensionalidade	tSNE, PCA, KPCA
Redes Neurais / Deep Learning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





Data Science
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science
Academy**

Como Selecionar o Algoritmo Ideal para Cada Problema?





Quando alguém perguntar a você:
Qual algoritmo de Machine Learning devo usar?

A resposta correta será:

Depende.

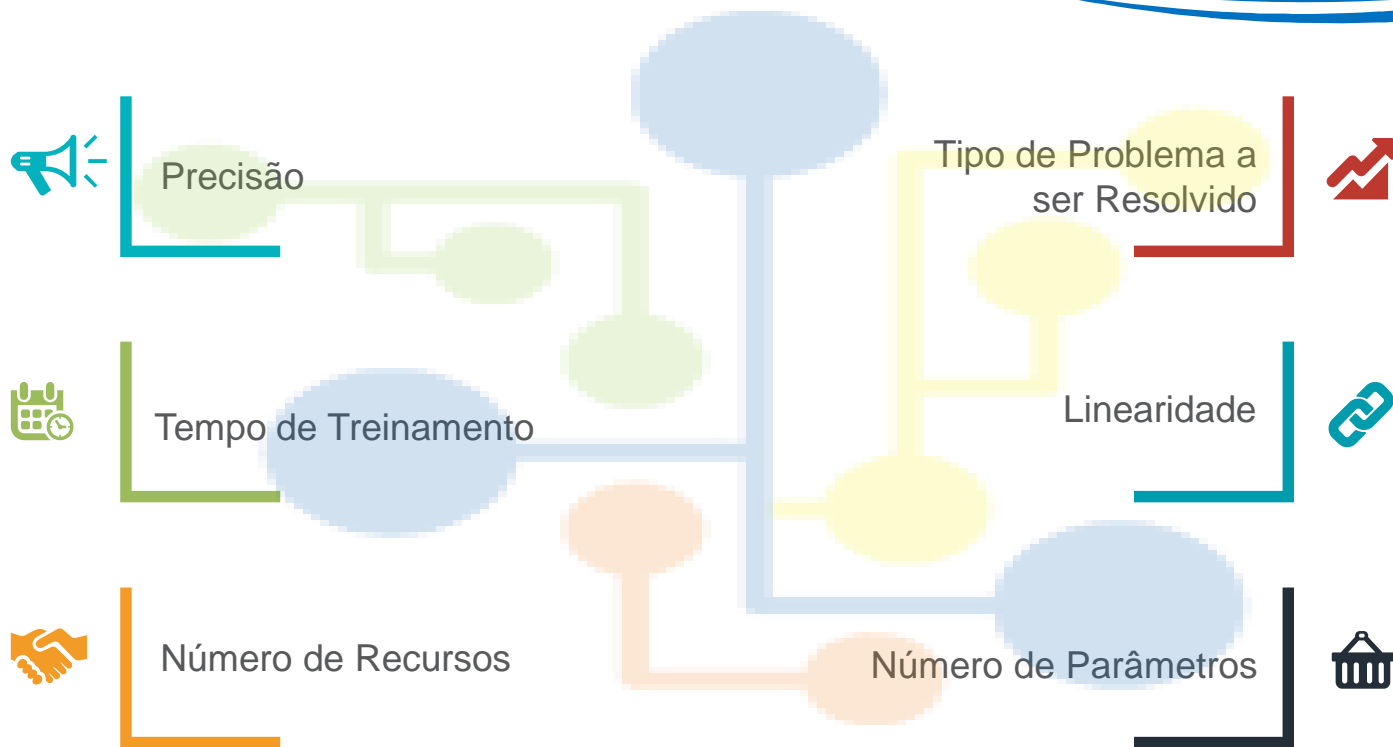


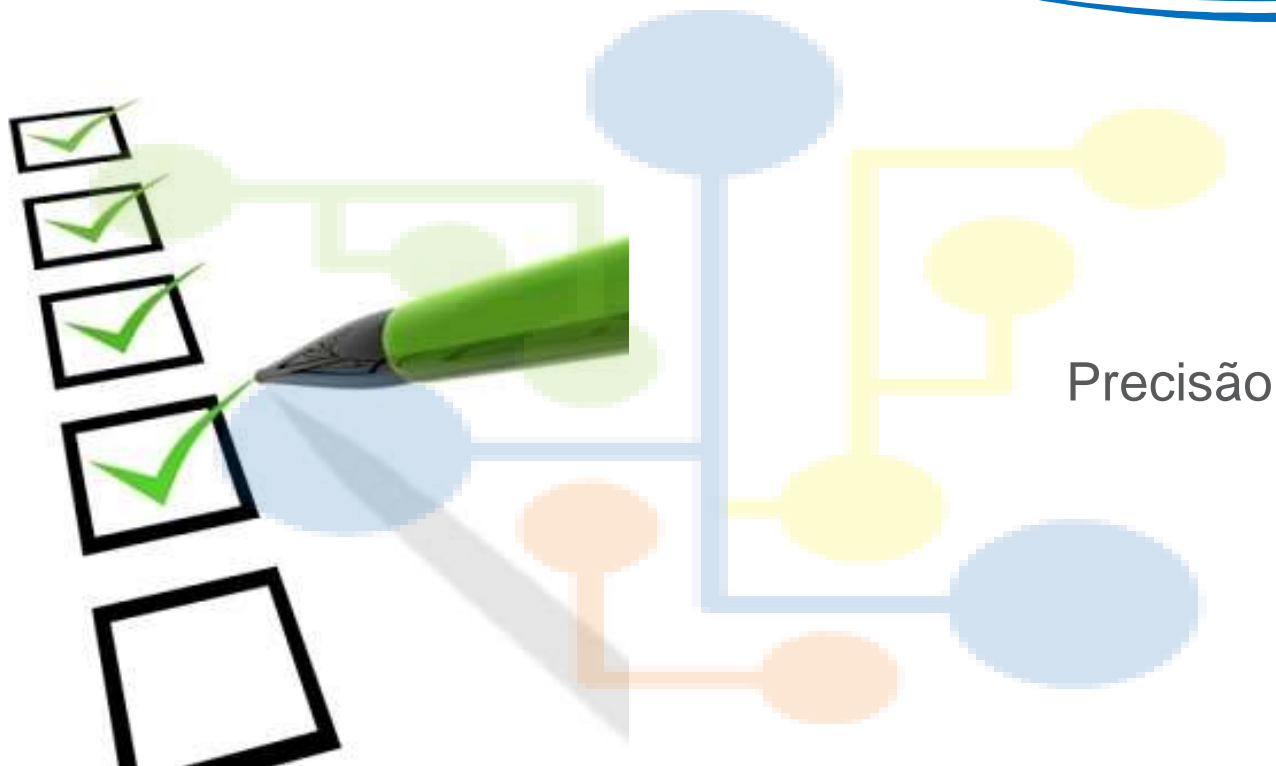


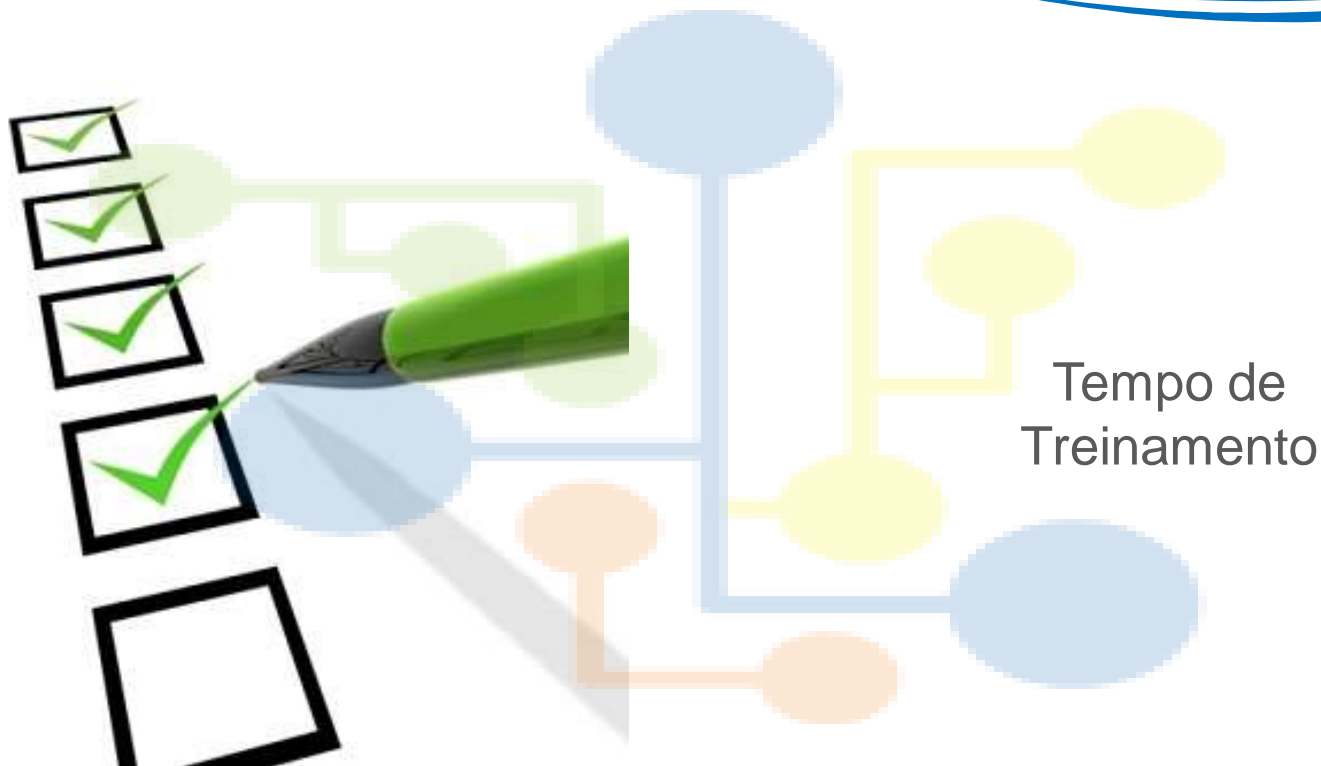
- Árvores de decisão
- Random Forests
- Descoberta de associações e sequência
- Boosting e bagging de gradiente
- Máquinas de vetores de suporte
- Redes neurais
- Mapeamento de nearest-neighbor
- Cluster k-means
- Mapas auto-organizáveis
- Técnicas de otimização de busca local (por ex., algoritmos genéticos)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

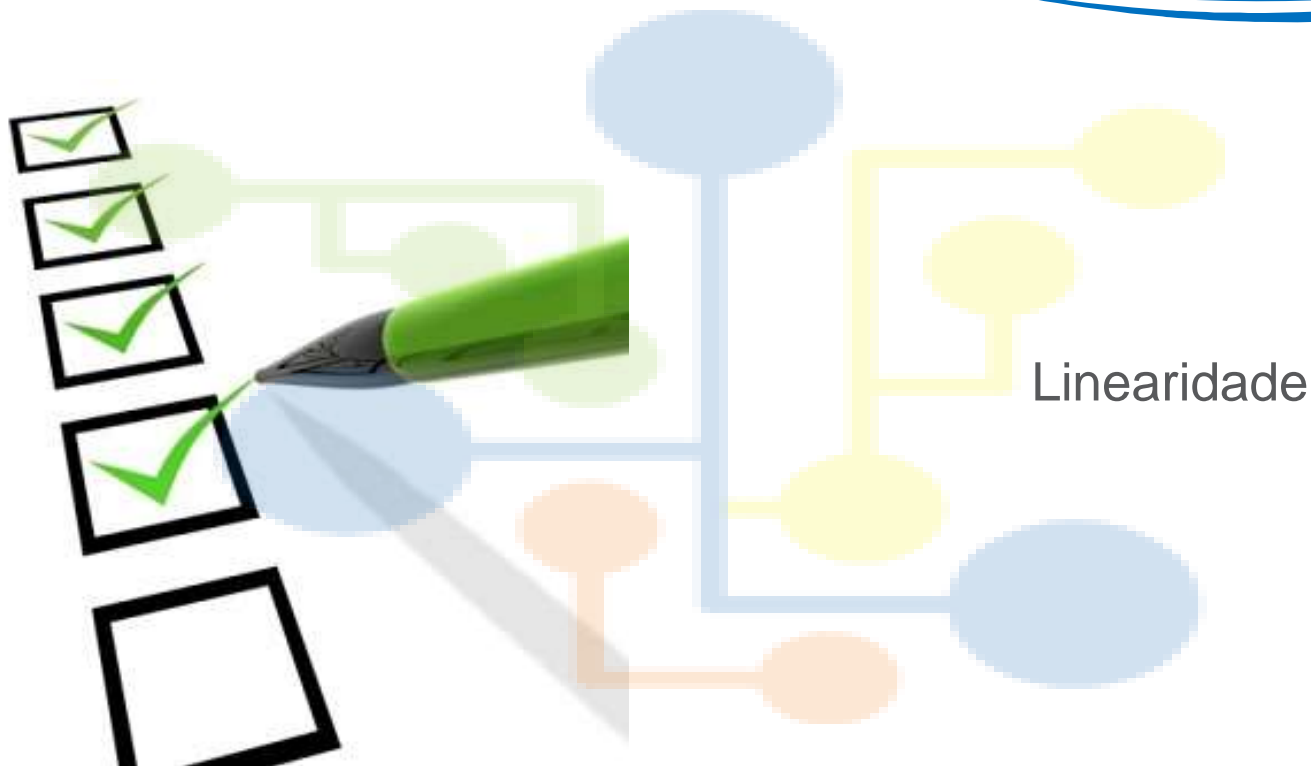
São muitos os algoritmos
de Machine Learning

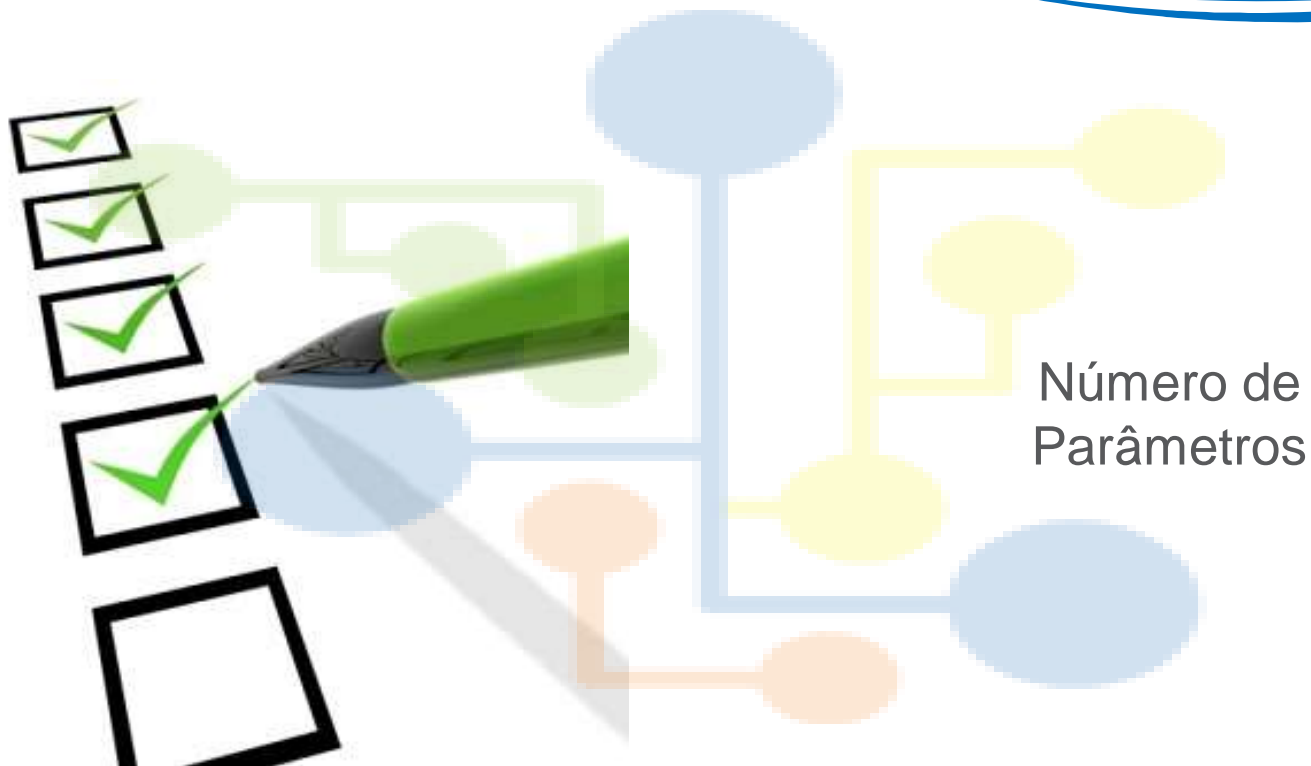


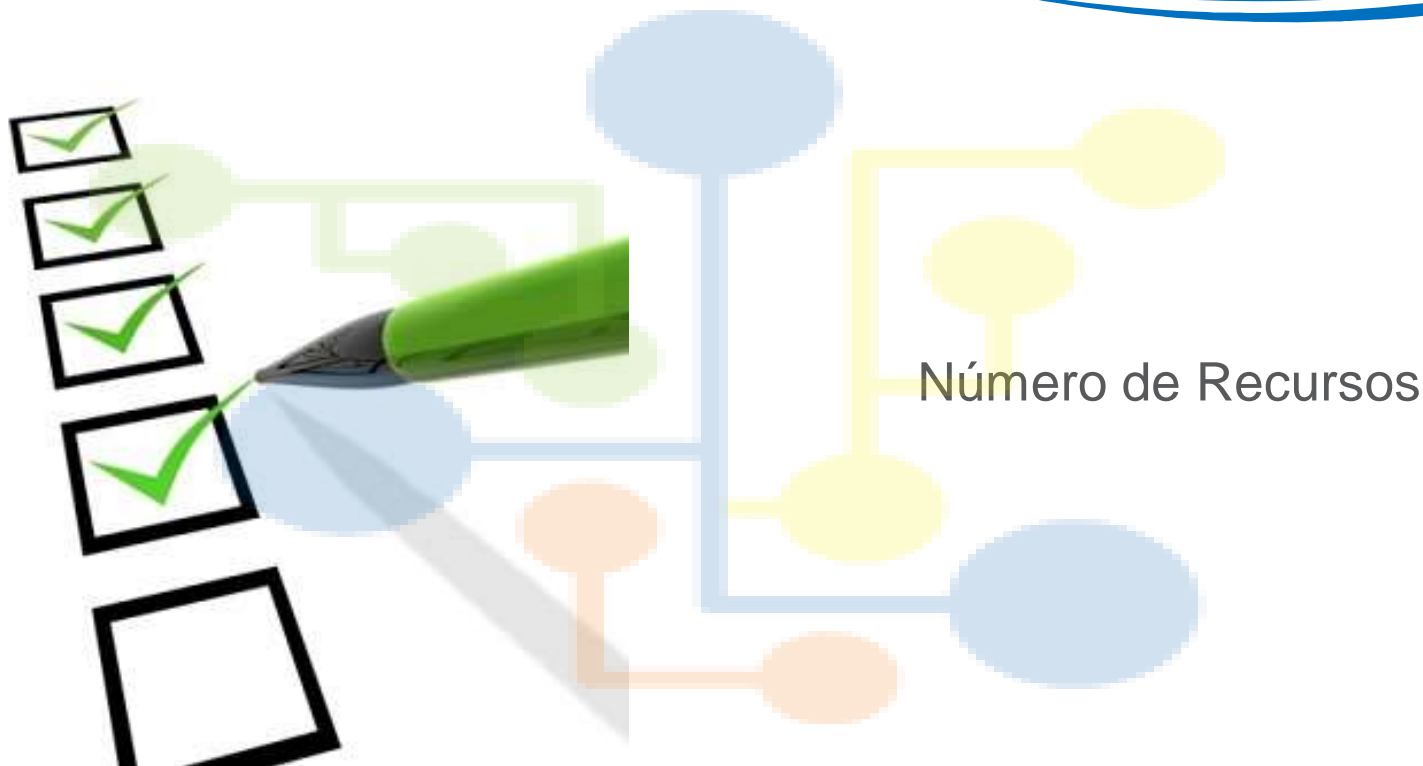














Comparação entre os principais algoritmos





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	■ Moderado	■ Moderado	■ Alto
Árvore de Decisão	■ Moderado	■ Alto	N/A
Random Forest	■ Alto	■ Alto	N/A
Redes Neurais	■ Alto	■ Alto	N/A
Poisson	■ Alto	■ Alto	■ Alto



Não Supervisionados

■ Alto
■ Moderado

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



Continue Trilhando uma Excelente Jornada de Aprendizagem!

Muito Obrigado!