



x

# MACHINE LEARNING

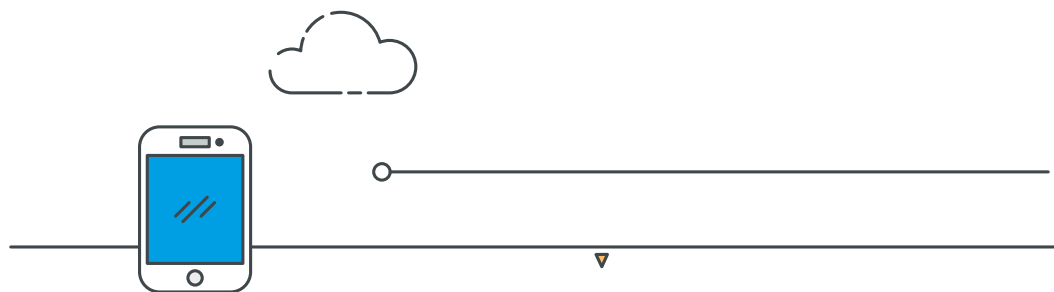


Machine Learning é um método de análise de dados que automatiza o desenvolvimento de modelos analíticos. Também chamada de Aprendizado de Máquina, é um tipo de Inteligência Artificial que capacita o computador a aprender de forma interativa a partir de dados. Isto permite que as máquinas encontrem informações sem serem necessariamente programadas para isso.

Hoje em dia vemos exemplos de Machine Learning ao utilizar as redes sociais – os feeds de algumas redes são personalizados a partir dos interesses dos usuários. A Machine Learning é um sub-campo da ciência da computação e baseia-se em análise estatística e análise preditiva para detectar padrões e encontrar dados.

Os algoritmos de aprendizagem automática podem fazer previsões sobre dados e aprender com os seus erros. Eles operam a partir de inputs de amostras para ter base para tomar tais decisões.

# O QUE É ?



O Machine Learning como campo de estudo se desenvolveu a partir dos esforços científicos para desenvolver ideias sobre inteligência artificial – alguns estudiosos se debruçavam sobre a possibilidade de fazer computadores aprenderem baseados em dados.

O Aprendizado de Máquina como campo de estudo só começou a existir realmente a partir dos anos 90. Seu desafio não era mais somente alcançar a inteligência artificial e sim selecionar problemas práticos com métodos da estatística e probabilidade. Isso também foi possível graças ao desenvolvimento da comunicação digital e internet. Hoje está cada vez mais acessível o aprendizado de máquina, pois o processamento computacional está mais barato e poderoso e o armazenamento de dados mais eficiente. É possível produzir modelos que analisam dados maiores e mais complexos e fornecem resultados mais rápidos e precisos em uma escala muito grande. Isto gera insights que levam a decisões e ações inteligentes que não contam com humanos para acontecer. O Machine Learning se baseia no raciocínio indutivo, que extrai regras e padrões de grandes conjuntos de dados.

As pessoas aprendem a partir de generalizações. E isso não é diferente na forma que o aprendizado foi conduzido com as máquinas, em que elas desenvolvem padrões a serem seguidos baseados em experimentações de algum conjunto de dados. Essa é uma teoria conhecida como aprendizagem computacional, do ramo da ciência da computação teórica.

A aprendizagem de máquinas é usada em muitas tarefas geridas por computadores onde criar e programar algoritmos explícitos é impraticável. Exemplos de aplicações incluem filtragem de spam, reconhecimento ótico de caracteres (OCR), processamento de linguagem natural, motores de busca, diagnósticos médicos e etc.

# ABORDAGENS

O aprendizado de máquinas é conduzido a partir da adoção de alguns métodos. Os dois mais importantes são o aprendizado supervisionado (responsável por 70% do mercado) e o aprendizado não supervisionado (responsável por 10% a 20%). Além desses dois existem o aprendizado semisupervisionado e o aprendizado por reforço.

## Aprendizado supervisionado

No aprendizado supervisionado a máquina aprende a partir de um modelo rotulado, com input e outputs desejadas. O algoritmo recebe esse modelo e aprende comparando seu exemplo com a realidade para encontrar erros e modificá-lo de acordo. O aprendizado supervisionado utiliza padrões para prever os valores dos rótulos em dados não rotulados usando métodos como a classificação, regressão, previsão e boosting do gradiente. Ele é muito usado em aplicações onde os dados históricos fornecem informações para prever prováveis acontecimentos futuros.

## Aprendizado não supervisionado

Nesse método o objetivo do sistema é explorar os dados e encontrar alguma estrutura que faça sentido neles. O aprendizado não supervisionado é usado quando não possuímos rótulos históricos – não tem uma resposta prévia certa.

As técnicas mais usadas são, entre outras, mapas auto-organizáveis, mapeamento do vizinho mais próximo, agrupamento k-means e decomposição em valores singulares.

A aprendizagem não supervisionada pode ser um objetivo em si mesma (descobrir novos padrões nos dados) ou um meio com um fim.

Os algoritmos desse método também são usados para segmentar tópicos de texto, recomendar itens e identificar os valores discrepantes dos dados.

## Aprendizado Semisupervisionado

O aprendizado semisupervisionado é usado nas mesmas situações que o supervisionado. A diferença é que ele usa tanto dados rotulados como os sem informações para seu aprendizado – uma pequena parte com rótulos e uma grande quantidade sem, pois os dados não marcados são mais baratos e não precisam de tanto esforço para serem adquiridos. Este método, então, é útil quando o custo associado à rotulagem é muito alto para permitir um processo de aprendizado totalmente supervisionado. Nesse tipo de treinamento podem ser usadas técnicas como classificação, regressão e previsão.

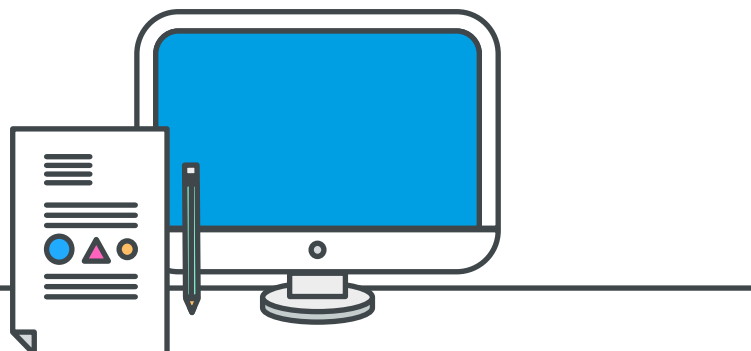
## Aprendizado por Reforço

Através deste método o algoritmo descobre pela tentativa e erro quais são as ações que geram as maiores recompensas.

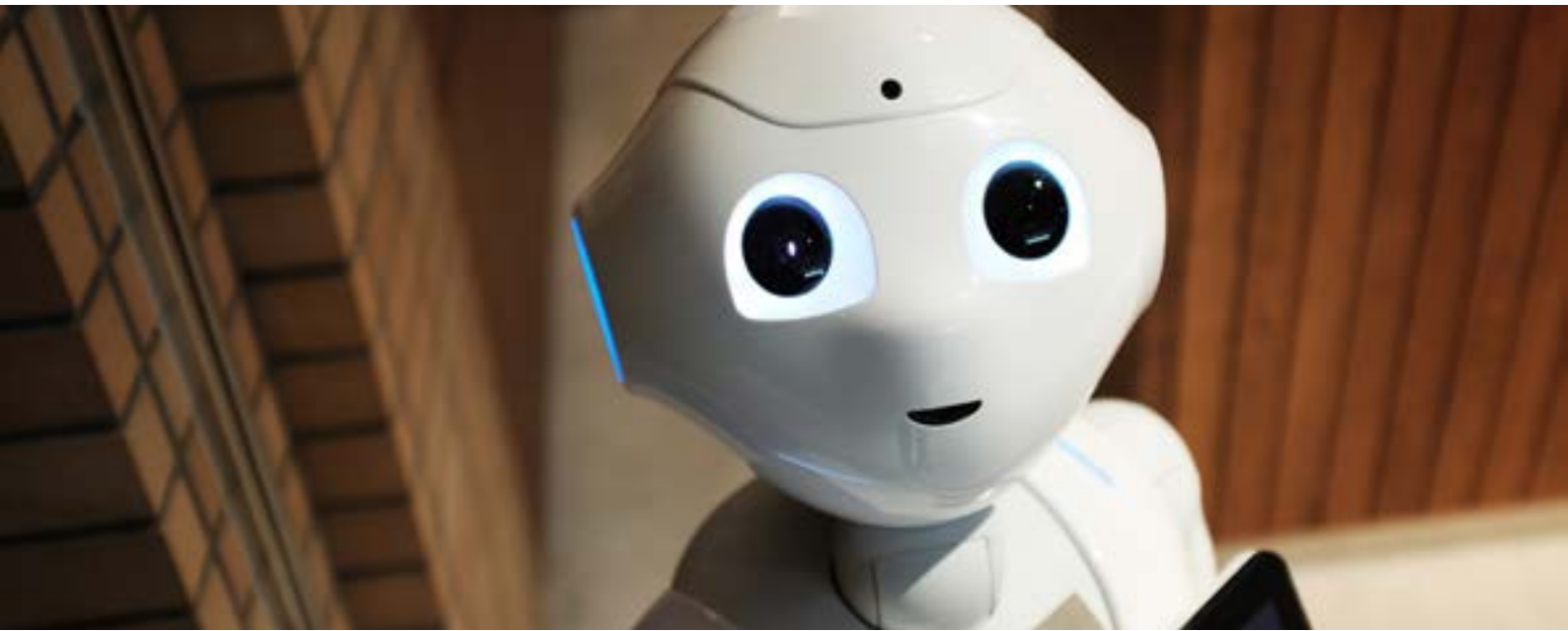
São três os itens que compõem o aprendizado por reforço – o agente (tomador de decisões), o ambiente (tudo com o que o agente interage) e as ações (o que pode ser feito pelo agente).

O objetivo é que o tomador de decisões escolha ações que maximizem a recompensa esperada em um período de tempo. O resultado será mais rápido seguindo uma boa política, então o objetivo final desse tipo de aprendizado é aprender a melhor política. Ou seja, aprender a jogar já jogando contra um oponente.

A aprendizagem por reforço é muito utilizada para navegação, jogos e robótica.



# MACHINE LEARNING NOS DIAS DE HOJE



Com o avanço das tecnologias, nos dias atuais o aprendizado de máquina não é o mesmo do passo. Enquanto antigamente a base era somente o algoritmo, hoje a análise de big data a partir de complexos cálculos está cada vez mais desenvolvida.

Exemplos como os carros autônomos do Google que dirigem sozinhos ou as recomendações do Netflix são cada vez mais comuns e desenvolvidos.

Os termos Mineração de Dados e Deep Learning são bastante presentes no nosso cotidiano.

## MINERAÇÃO DE DADOS

A mineração de dados se aproxima do cérebro humano no sentido de observar padrões e formular hipóteses para as informações. No momento que os dados se tornaram muito volumosos, a técnica entra instrumentalizando o computador para funcionar dessa forma e encontrar relações razoáveis.

Também chamada de prospecção de dados, mineração é a exploração de grandes quantidades de dados evidenciando padrões que sejam consistentes – regras de associação ou sequências temporais

que detectam relacionamentos sistemáticos entre as variáveis. Essa exploração é baseada em algoritmos, redes neurais ou estatística. A informação pode ser apresentada de diversas formas: agrupamentos, hipóteses, regras, entre outras.

A mineração de dados pode ser utilizada desde pesquisas científicas até estratégias empresariais. Várias áreas do conhecimento se interlaçam para construir o processo de análise e através dele é possível extrair informações importantes de comportamento, sugerir tendências e permitir melhores ações.

## DEEP LEARNING

Deep Learning (ou aprendizagem profunda, em português) se refere ao desenvolvimento mais aprofundado da machine learning. Combina avanços tecnológicos com tipos especiais de redes neurais para aprender a analisar padrões mais complicados em uma grande quantidade de dados. A aprendizagem profunda lida diretamente com reconhecimento de padrões complexos, como objetos em imagens ou palavras em sons e são usados para, por exemplo, tradução automática de idiomas ou diagnósticos médicos. Assim como o cérebro humano pode identificar um objeto em milissegundos, o deep learning pode realizar isto com quase a mesma velocidade e precisão.

A aprendizagem profunda tem a capacidade de se adaptar a diferentes variantes e obter informações a partir de dados parciais. É o mais próximo da inteligência humana que as máquinas conseguiram chegar.

# CONTRAS

O aprendizado de máquinas dá às instituições o potencial de tomar decisões mais apuradas, baseadas em dados, para resolver problemas. Porém ela não é mágica –apresenta a maioria dos desafios que outros métodos de análise apresentam. Na maioria do tempo os cientistas de dados estão ocupados preparando as informações e lidando com problemas de qualidade. Qualidade de dados é essencial para conseguir resultados mais apurados a partir das análises. Alguns sistemas ditos confiáveis podem falhar completamente no fornecimento e análise dos dados necessários para a organização. Isso acontece em uma escala muito menor que os erros humanos, mas ainda não é um sistema que pode se dizer perfeito.

## ONDE APRENDER MAIS

**Coursera**– Com professor de Stanford. Gratuito, porém sem emissão de certificado. Conta com versão paga que oferece algumas funcionalidades atuais que incluem a emissão.

**Udacity + Google**– Curso de Deep Learning oferecido pelo Google através da Udacity. *Gratuito*.

**EDX** –Para quem domina o inglês. Conta com uma equipe de professores da Universidade de Columbia. Gratuito e com a versão paga com emissão de certificado



