## 1. Introdução ao Aprendizado de Máquina

Nesta aula você vai aprender:

* O que é o Aprendizado de Máquina e o novo paradigma de programação que ele traz
* A importância e as aplicações do Aprendizado de Máquina
* Alguns conceitos e termos básicos que envolvem o Aprendizado de Máquina

Você certamente já teve contato com o Aprendizado de Máquina e talvez nem tenha se dado conta disso. É o Aprendizado de Máquina que está por traz dos feeds das suas redes sociais, dos anúncios do Instagram e das sugestões de filmes, músicas e compras da Netflix, Spotify e Amazon. São também algoritmos de Aprendizado de Máquina que estão por traz das transcrições e traduções automáticas de texto do Google Translator ou do YouTube e, se você já teve a experiência de um cartão bloqueado por uma suspeita de fraude sem que você tenha solicitado o bloqueio, certamente isso envolveu o Aprendizado de Máquina. Ele também alimenta chatbots e veículos autônomas, facilita o diagnóstico médico a partir de imagens, fornece preços dinâmicos para corridas de serviços como Uber, prevê falhas de uma planta industrial que envia informações de seus equipamentos por sensores e busca melhores rotas para os motoristas no Waze.

Algumas pessoas usam os termos Inteligência Artificial e Aprendizado de Máquina quase como sinônimos. Mas o mais provável quando vemos empresas fazendo uso de Inteligência Artificial, é que estejam mesmo empregando o Aprendizado de Máquina que é um subcampo da IA que permite aos computadores aprender e realizar tarefas sem serem explicitamente programados. Em 2020, mais de 65% das empresas já fazem do Aprendizado de Máquina e a quase a totalidade das empresas acredita que ele será um fator crítico de sucesso para os negócios e, embora nem todos precisem conhecer os detalhes de como isso é feito, todos precisaremos saber o que Aprendizado de Máquina pode e não pode fazer por nós, e as implicações dessa enorme transformação.

# Aprendizado de Máquina: Um Novo Paradigma

O Aprendizado de Máquina, ou *Machine Learning* (ML), é um subcampo da Inteligência Artificial. A IA busca criar sistemas capazes de imitar o comportamento humano inteligente o que pode ser feito de várias formas.

O uso da lógica formal como acima, ou um circuito autômato finito que controla a porta automática em um shopping, podem reproduzir comportamentos inteligentes. Eles fazem parte do estudo da IA, mas não são parte do Aprendizado de Máquina. Há também uma série de outros termos correlatos e as Ciência de Dados e Deep Learning (Redes Neurais Profundas) são áreas de comumente de interceptam e que, embora não haja fronteiras nitidamente definidas, não são a mesma coisa.

Linha do tempo

Descrição gerada automaticamente

**Figura 1. Diferentes Áreas relacionadas ao Aprendizado de Máquina.**

Assim o Aprendizado de Máquina é uma, entre muitas outra maneiras, de empregarmos a IA. Há muitas definições para o Aprendizado de Máquina e podemos empregar por exemplo a definição dada pelo pioneiro de jogos e IA, ​​Arthur Samuel (1901-1990):

**O Aprendizado de Máquina é o campo de estudo que dá aos computadores a capacidade de aprender sem serem explicitamente programados**

Essa definição traz um aspecto fundamental do Aprendizado de Máquina, que cria um novo paradigma de como *programamos*.

Diagrama

Descrição gerada automaticamente

**Figura 2. O Aprendizado de Máquina: cria um novo paradigma de**

**como programamos os computadores.**

Você certamente não terá dificuldade de criar um programa para converter temperaturas de (Celsius) para (Fahrenheit).

def Celsius2Fahrenheit(Celsius):  
 return Celsius \* 9/5 + 32

print( Celsius2Fahrenheit(0) )  
print( Celsius2Fahrenheit(27) )

32.0  
80.6

Mas será que um 'programa' pode criar esse programa de conversão de temperaturas? A ideia do Aprendizado de Máquina é que sim. Mas para que seja possível o aprendizado precisamos partir de algum ponto que são os dados.

import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
%matplotlib inline  
import seaborn as sns

df = pd.DataFrame()  
df['celsius'] = np.arange(-20,60)  
df['fahrenheit'] = Celsius2Fahrenheit( df['celsius'] )  
df.head()

celsius fahrenheit  
0 -20 -4.0  
1 -19 -2.2  
2 -18 -0.4  
3 -17 1.4  
4 -16 3.2

Se tivermos um conjunto de dados com valores de temperaturas em (Celsius) e seus respectivos valores em (Fahrenheit), podemos criar um programa que *aprende* esse padrão e empregá-lo para a conversão de novos dados.

from sklearn.linear\_model import LinearRegression  
  
X = df[['celsius']]   
y = df['fahrenheit']  
  
regressor = LinearRegression()  
regressor.fit(X, y)

LinearRegression()

O programa acima cria um *modelo* (um modelo linear) para os dados. Esse modelo busca capturar o *padrão* dos dados, isto é, a regra que leva os valores de (Celsius) (Fahrenheit), e cria um *programa* com essa regra que pode ser aplicado a novos valores.

Fahrenheit\_prediction = regressor.predict(pd.DataFrame({'celsius':[0,27]}))  
print(Fahrenheit\_prediction)

[32. 80.6]

Agora, nosso novo programa regressor, converte as temperaturas da mesma forma que o programa Celsius2Fahrenheit onde explicitamente codificamos a regra de conversão de Celsius para Fahrenheit, mas *criamos* esse programa unicamente a partir de dados, sem conhecermos explicitamente a regra de conversão.

# The coefficients  
print("Coefficients: \n", regressor.coef\_, regressor.intercept\_)

Coefficients:   
 [1.8] 32.0

Nosso regressor é um modelo de regressão linear, um dos modelos de Aprendizado de Máquina mais simples, e que mapeia dados de entrada e saída na forma de uma equação linear.

O *aprendizado* consiste em encontrar, nesse modelo, os coeficientes que melhor se ajustam aos dados e você pode notar que os coeficientes obtidos são exatamente os mesmos da nossa fórmula colocada explicitamente no nosso código Celsius2Fahrenheit.

Talvez você ache pouco interessante o exemplo que apresentamos. Afinal parece ter sido muito mais fácil criar o programa Celsius2Fahrenheit e, de fato, quando conhecemos a regra por traz dessas transformações é muito mais simples explicitarmos isso em um programa que empregarmos o Aprendizado de Máquina. Um modelo conhecido (a regra explícita) apresentará sempre um resultado melhor. Mas e quando você não conhece essa regra?

Há muitos problemas práticos em que não conhecemos essa regra (ela pode até mesmo não existir) e não podemos, portanto, explicitá-la em um programa. É nesses casos que o Aprendizado de Máquina será útil.

Suponha agora que você não saiba como converter temperaturas (Fahrenheit) para (Celsius). É claro que você pode pegar a fórmula anterior e *deduzir*, mas imagine por um momento que você não tem como conhecer e explicitar essa regra (e de fato, diferentemente da fórmula você não encontrará essa fórmula aqui no texto!). Podemos aprender essa regra apenas alterando nosso programa de Aprendizado de Máquina, trocando suas entradas e saídas,

from sklearn.linear\_model import LinearRegression  
  
X = df[['fahrenheit']]  
y = df['celsius']   
  
regressor = LinearRegression()  
regressor.fit(X, y)

LinearRegression()

Celsius\_prediction = regressor.predict(pd.DataFrame({'fahrenheit':[32,80.6]}))  
print(Celsius\_prediction)

[ 0. 27.]

O mesmo programa regressor pode ser empregado para outras transformações de temperatura, como de Kelvin Fahrenheit Celsius, bastando para isso termos essas respectivas temperaturas.

df['kelvin'] = (df['fahrenheit'] - 32) \* 5/9 + 273.15

from sklearn.linear\_model import LinearRegression  
  
X = df[['celsius']]  
y = df['kelvin']   
  
regressor = LinearRegression()  
regressor.fit(X, y)  
  
Kelvin\_prediction = regressor.predict(pd.DataFrame({'celsius':[27,0]}))  
print(Kelvin\_prediction)

[300.15 273.15]

O valor aqui é aproximado, pois temos poucos valores de temperaturas Kelvin em nosso conjunto.

Esse é um exemplo simples, mas muitos outros problemas, como a estimativa do preço de imóveis a partir de suas características (área, idade do imóvel, localização etc.) ou o tempo de recuperação de um paciente a partir de dados de seu exame, em que não conhecemos a regra explícita (e podem mesmo não haver) para a formação de preços e do tempo para o reestabelecimento do paciente, podem ser resolvidos de modo bastante satisfatório do mesmo modo.

**O Aprendizado de Máquina se aplica a problemas em que não conhecemos as regras explícitas que mapeiam as entradas e saídas desejadas. Se a regra é conhecida o modelo explícito será sempre será mais simples e terá melhores resultados.**

# Um pouco de Terminologia

## Rótulos ou Variável Objetivo,

Um rótulo ou variável objetivo é a saída que estamos buscando. No nosso exemplo buscamos temperaturas, mas poderíamos estar buscando o preço de imóveis ou tempo de recuperação de pacientes acometidos de uma doença. Muitas vezes essa saída assume valores discretos, comumente chamadas de **classes**, como os valores 'positivo' e 'negativo' para o resultado de um exame de diagnóstico ou detecção de *spam* na caixa de e-mails, ou **grupos**, como 'política', 'economia' e 'esportes' na categorização de notícias e documentos.

## Recursos, Features ou Atributos,

Um recurso, feature ou atributos, é uma variável de entrada que pode ter uma ou mais dimensões e corresponde a características daquela instância e que, imaginamos, possam dizer algo sobre o nosso objetivo . Esses atributos podem ser dos mais variados dependendo do problema, de imagens de ressonância para detecção de tumor, a quantidade de palavras e número de letras maiúsculas em mensagens para a detecção de *spam*, até a hora e valores das transações para a detecção de fraudes em cartões de crédito.

## Conjunto de Treinamento

Em quaisquer casos o Aprendizado de Máquina está sempre associado a um conjunto de dados que são empregados no aprendizado. Esses dados quando apresentam um rótulo, ou variável objetivo, que desejamos conhecer, são frequentemente chamados de **conjunto de treinamento**. Há, entretanto, no Aprendizado de Máquina situações em que empregamos conjuntos dados não rotulados e você irá aprender mais sobre isso adiante.

## Modelos

Um **modelo** define a relação entre os recursos e o rótulo, isto é, entre as entradas e as saídas (não trataremos por hora os casos em que os dados não são rotulados). Por exemplo, um modelo de Regressão semelhante ao que empregamos antes, pode ser utilizado para se obter uma estimativa de preço de um imóvel a partir de seus atributos como área, idade do imóvel, localização etc.

O **treinamento** ou **aprendizado**, significa criar ou treinar o modelo que melhor se ajusta aos dados. No nosso exemplo de temperaturas empregamos *um* modelo linear e o treinamento ou aprendizado consistiu em obter os coeficientes que melhor ajustavam *o* modelo aos dados (não se preocupe, você vai aprender como isso é feito mais adiante). Em um modelo para predição de valores de imóveis a partir de um modelo linear o aprendizado seria obter os coeficientes do modelo,

Com os valores dos coeficientes obtidos obtemos *o* modelo, dentro de todos os modelos lineares possíveis, e podemos empregá-lo na estimativa, predição ou **inferência** de novos preços.

# Tipos de Aprendizado de Máquina

Os sistemas de Aprendizado de Máquina nem sempre são empregados para inferência e, embora aqui nos dediquemos unicamente a esse tipo de aprendizado (o Aprendizado Supervisionado), é importante você entender que há outras formas de aprendizado e que você aprenderá nas disciplinas subsequentes.

Sistemas de Aprendizado de Máquina podem simplesmente ser **descritivos**, o que significa que o sistema emprega os dados para explicar o que aconteceu. Por exemplo, podemos *segmentar* um conjunto de consumidores com base em suas características (idade, perfil de consumo etc.) e, com isso, direcionar uma campanha de marketing. Você não faz aqui qualquer inferência. Eles são **preditivos**, quando o sistema usa os dados para prever o que pode acontecer. É o caso dos exemplos que vimos antes: você prevê que uma certa transação de cartão de crédito com base em um conjunto de exemplos de outras transações com valor, local, horário, uso do cartão nas últimas 24h etc. é com uma certa probabilidade uma fraude (e neste caso impede que isso ocorra!). E há ainda os sistemas **prescritivos**, em que o sistema usará os dados para fazer sugestões sobre quais ações tomar, como no caso de um robô-aspirador que vai tomando ações no ambiente e se tornando mais eficiente à medida que *aprende*, isto é, à medida que obtêm mais dados sobre o ambiente. Os sistemas de Aprendizado de Máquina são assim divididos em três subcategorias,

* + **Aprendizado Supervisionado**
  + **Aprendizado Não Supervisionado**
  + **Aprendizado de Reforço**

Havendo ainda sistemas híbridos que combinam descrição, predição e precrição.

Diagrama

Descrição gerada automaticamente

**Figura 3. Aprendizado de Máquina: Tipos, Tarefas e alguns algoritmos.**

# Aprendizado Supervisionado

O aprendizado supervisionado é uma abordagem de aprendizado de máquina definida pelo uso de conjuntos de dados **rotulados**. Por exemplo você pode ter dados de transações de crédito que são previamente rotulados como fraude ou não fraude, dados de exames de tumores rotulados como malignos ou benignos, ou ainda dados de imóveis *rotulados* com os seus preços. Esses dados funcionam como um conjunto de exemplos para o aprendizado e são empregados para treinar ou *supervisionar* os algoritmos para classificar novos dados ou prever valores.

Em geral o aprendizado supervisionado trata de dois tipos de problemas, ou tarefas:

* + **Regressão**
  + **Classificação**

Modelos de **regressão** prevêm valores contínuos. Por exemplo, os preços de imóveis ou a probabilidade de um usuário clicar em um anúncio. Modelos de **classificação** preveem valores discretos, geralmente denominados como **classes**. Por exemplo, determinar se uma mensagem de e-mail é spam ou não é spam, se uma foto é de paisagem, uma pessoa ou um objeto, ou se uma transação de crédito é fraude ou não.

Tabela

Descrição gerada automaticamente

**Figura 4. Acima um mesmo conjunto de dados empregado para o aprendizado supervisionado em problemas de classificação e regressão. são os rótulos associados a cada instância de dados com variáveis preditoras . Os modelos supervisionados criam, a partir dos exemplos do conjunto de treinamento, uma função que reproduz do mesmo padrão de mapeamento das entradas e saídas do conjunto de treinamento. A classificação trata da predição de valores discretos (as classes diagnosis à esquerda, com valores M(aligno) ou B(enigno)). A regressão trata da predição de valores contínuos, como a área média do tumor (à direita).**

# Aprendizado não Supervisionado

O aprendizado não supervisionado usa algoritmos de aprendizado de máquina para analisar e agrupar conjuntos de dados não rotulados. São modelos descritivos, mais analíticos, que os modelos supervisionados e não buscam fornecer uma predição de valores ou categorias. A ideia desses algoritmos é encontrar padrões ocultos nos dados sem a necessidade de qualquer intervenção humana ou *rótulos* pré-definidos e, por isso, denominados como *não supervisionados*.

Modelos de aprendizagem não supervisionados são empregados vários tipos de tarefas ou problemas e os principais são: **agrupamentos** (ou **clusterização**), **associações**, **redução de dimensionalidade** e **detecção de anomalias**.

Os modelos de **Clustering** permitem agrupar dados não rotulados com base em suas semelhanças ou diferenças e são empregados em problemas de segmentação de clientes, produtos etc. Algoritmos de **Associação**, podem ser usados ​​para análise de cesta de compras ou sistemas de recomendação e permitem identificar que, clientes que compram um determinado item, também são propensos a comprar outro. Algoritmos como PCA (*Principal Component Analysis*), permitem **reduzir a dimensionalidade** de dados centenas ou milhares de atributos para facilitar uma análise dos dados. E modelos de **detecção de anomalia**, permitem identificar um comportamento suspeito de tráfego de internet em um site, indicando uma falha do sistema ou uma tentativa de *deny-of-service*, ou ainda detectar um movimento em uma câmera de segurança.

Diagrama

Descrição gerada automaticamente

**Figura 5. Exemplos de Aprendizado Supervisionado. Você mesmo, apenas olhando as figuras, poderia identificar tanto os agrupamentos como as anomalias sem, entretanto, ter qualquer exemplo ou rótulos indicando esses grupos e esses pontos atípicos nos dados.**

# Aprendizado de Reforço

O **aprendizado de reforço** se utiliza de um sistema com um fator de recompensa para o aprendizado. O treinamento ocorre por meio de tentativa e erro estabelecendo-se um sistema de recompensa. Um agente toma uma sequência de decisões na busca de atingir uma meta (o problema que se deseja resolver). As ações que resultam em se aproximar da meta são recompensadas, e ações que se afastam da meta são penalizadas. O agente, desse modo, vai *aprendendo* e se aproximando da solução do problema ao maximizar a recompensa total que pode ser, por exemplo, obter um máximo de pontos em um jogo. Esse tipo de aprendizado é empregado principalmente em robótica e robôs-aspirador implementam modelos simples desse tipo de aprendizado.

# Usos do Aprendizado de Máquina

Ao longo do texto demos vários exemplos de aplicações que vem sendo feitas do Aprendizado de Máquina. Para algumas empresas o Aprendizado de Máquina está no centro de seus negócios. É caso das redes sociais, do Facebook e o Instagram, ao LinkedIn, ou a Netflix e o sistema de busca da Google. Mas de empresas do varejo à empresas de extração, saúde ou financeiras, todos os setores e mesmo a Ciência, estão adotando o Aprendizado de Máquina, mesmo que não sendo esse o seu propósito central. Revisamos aqui alguns dos usos mais comuns que encontram-se espalhados ao longo do texto para ilustrar aspectos específicos do Aprendizado de Máquina.

* **Sistemas de recomendação**. Recomendações da Netflix, YouTube ou Instagram, de mensagens do Twitter ou de anúncios no FaceBook que queremos ver buscam extrair o padrão de preferência dos usuários.
* **Detecção de fraude**. Identificação de transações de cartão de crédito ou compras fraudulentas em sites de e-commerce, tentativas fraudulentas de login etc.
* **Chatbots**. Chatbots online em que os clientes, interagem por texto ou mesmo fala, com uma máquina que guia o usuário para um determinado resultado empregam uma série de recursos de Aprendizado de Máquina, do processamento de linguagem natural ao aprendizado de reforço.
* **Diagnóstico médico**. Algoritmos de aprendizado de máquina são empregados para examinar imagens médicas ou outros dados de pacientes e auxiliar no diagnóstico precoce de doenças como câncer ou diabetes.
* **Veículos Autônomos**. Empregam um grande número de recursos de aprendizado de máquina para segmentação de cenas (identificação de ruas, calçadas, prédios e pessoas em uma cena), movimentação autônoma, predição e tratamento de colisão etc. em geral com emprego de algoritmos de Deep Learning.
* **Análise e Classificação de Risco**. Para precificação de crédito e seguros.

Placa azul com letras brancas

Descrição gerada automaticamente com confiança média

[**Machine Learning and Data Science Applications in Industry**](https://github.com/ashishpatel26/Real-time-ML-Project) **|** [**https://github.com/ashishpatel26/Real-time-ML-Project**](https://github.com/ashishpatel26/Real-time-ML-Project)

|  |  |  |
| --- | --- | --- |
|  |  |  |
| [Accommodation & Food](#accommodation) | [Agriculture](#agriculture) | [Banking & Insurance](#bankfin) |
| [Biotechnological & Life Sciences](#biotech) | [Construction & Engineering](#construction) | [Education & Research](#education) |
| [Emergency & Relief](#emergency) | [Finance](#finance) | [Manufacturing](#manufacturing) |
| [Government and Public Works](#public) | [Healthcare](#healthcare) | [Media & Publishing](#media) |
| [Justice, Law and Regulations](#legal) | [Miscellaneous](#miscellaneous) | [Accounting](#accounting) |
| [Real Estate, Rental & Leasing](#realestate) | [Utilities](#utilities) | [Wholesale & Retail](#wholesale) |

O site acima dá uma ideia das várias áreas de aplicação e traz detalhes dos diversos *cases* em cada segmento. Você pode explorar e focar em apenas alguns desses casos que forem mais do seu interesse.



**Figura 6. Carro Autônomo com Tecnologia Nacional.**

**Fonte:** [**https://revistapesquisa.fapesp.br/o-futuro-sem-motorista/**](https://revistapesquisa.fapesp.br/o-futuro-sem-motorista/)

Mapa

Descrição gerada automaticamente

**Figura 7. eVTOL aeronave elétrica de decolagem e pouso vertical**

**sendo desenvolvido pela EMBRAER-UBER.**

**Fonte:** [**https://revistapesquisa.fapesp.br/2019/12/03/a-chegada-dos-carros-voadores/**](https://revistapesquisa.fapesp.br/2019/12/03/a-chegada-dos-carros-voadores/)

Diagrama

Descrição gerada automaticamente

**Figura 8. Sistema Automático de identificação de lâminas de madeira para uso industrial.**

**Fonte:** [**https://revistapesquisa.fapesp.br/2017/07/18/identificacao-de-madeiras/**](https://revistapesquisa.fapesp.br/2017/07/18/identificacao-de-madeiras/)

Ponte de metal sobre água com cidade ao fundo

Descrição gerada automaticamente com confiança média

**Figura 9. Sistema Híbrido combina Mecânica dos Fluídos e Aprendizado de Máquina para exploração mais eficaz do Pré-Sal pela Petrobrás.**

**Fonte:** [**https://revistapesquisa.fapesp.br/producao-mais-eficaz-no-pre-sal/**](https://revistapesquisa.fapesp.br/producao-mais-eficaz-no-pre-sal/)

Imagem de jogo de vídeo game

Descrição gerada automaticamente com confiança média

**Figura 10. Mina de alumínio usa sistema automatizado**

**para tratar água de rejeito.**

**Fonte:** [**https://revistapesquisa.fapesp.br/mineracao-digital/**](https://revistapesquisa.fapesp.br/mineracao-digital/)

Padrão do plano de fundo

Descrição gerada automaticamente com confiança baixa

**Figura 11. Aplicações em Astronomia e Astrofísica:**

**Ciência no Aprendizado de Máquina. Fonte:** [**https://amazon.com/**](https://amazon.com/)**.**

# Processo de Ciência de Dados X Aprendizado de Máquina

O Aprendizado de Máquina permite extrair conhecimento útil dos dados para resolver problemas práticos, mas não é um processo simples. Por isso, em geral, a construção de modelos envolve uma série de etapas que envolvem tarefas como de Análise e Exploração dos Dados, Preparação dos Dados, Distribuição e Monitoramento dos Sistemas etc. que você já viu ou estará aprendendo em outros módulos. O processo **CRISP-DM, Cross Industry Standard Process for Data Mining** (Processo Padrão Inter-Indústrias para Mineração de Dados) é um processo de fases bastante aceito na indústria para representar um ciclo completo de Ciência, Análise de Dados e construção de Modelos de Aprendizado de Máquina na Indústria e em geral seguido com poucas variações.

Diagrama

Descrição gerada automaticamente com confiança média

**Figura 12. As Fases do Processo CRISP-DM.**

**Fonte:** <http://meusite.mackenzie.br/rogerio/TIC2021S1/T1/crisp_visualguide.pdf>

Você consultar a documentação original desse modelo de referência aqui.

* [CRISP Visual Guide](http://meusite.mackenzie.br/rogerio/TIC2021S1/T1/crisp_visualguide.pdf) | Acesse aqui <http://meusite.mackenzie.br/rogerio/TIC2021S1/T1/crisp_visualguide.pdf>
* [CRISP DM 1.0](http://meusite.mackenzie.br/rogerio/TIC2021S1$/T1/crisp-dm.pdf) | Acesse aqui <http://meusite.mackenzie.br/rogerio/TIC2021S1$/T1/crisp-dm.pdf>

De qualquer modo, você encontra a seguir, uma descrição resumida de cada uma dessas fases.

### Entendimento do negócio

Esta fase inicial concentra-se em compreender os **objetivos e requisitos do projeto ou problema** de uma perspectiva de negócios e traça um plano preliminar do projeto.

### Entendimento dos dados

Envolve as principais atividades da **EDA**, **Análise Exploratória de Dados**. Dentre outras tarefas, verifica-se aqui a qualidade dos dados.

### Preparação de dados

A fase de preparação de dados inclui todas as atividades necessárias para tornar os dados adequados para a construção dos modelos. Envolvem a seleção de dados e atributos, transformações como normalização e recodificação dos dados, limpeza dos dados e tratamento de dados faltantes, combinações de dados etc.

**Importante destacar. Na prática, 80% do esforço de se criar modelos de Aprendizado de Máquina nas empresas envolve atividades preliminares de Análise, Exploração e Preparação dos Dados, consideradas bastante enfadonhas. Apenas 20%, ou menos, é dedicado à atividades de criação e análise dos algoritmos.**

### Modelagem

Nesta fase várias técnicas de modelagem são selecionadas e aplicadas, e seus parâmetros são ajustados na busca de melhores modelos. Algumas técnicas têm diferentes requisitos na preparação dos dados. Portanto, voltar à fase de preparação de dados é muitas vezes necessário.

**As fases não são completamente estanques e, frequentemente, identificamos em uma fase a necessidade de retornar a fase anterior.**

### Avaliação

Você construiu um modelo, ou vários modelos, mas é necessário avaliar a sua qualidade ou comparar diversos modelos que competem para atingir os objetivos do projeto. Aqui o papel das **métricas** de eficiência dos modelos é fundamental havendo diferentes métricas para diversos modelos de aprendizado.

### Distribuição

Concluído um modelo é necessário que ele seja organizado e distribuído para uso. Essa etapa na indústria, em geral, envolve a implementação de *pipelines* do modelo.

# Sumário da Aula

Nesta unidade você tomou contato os principais conceitos Aprendizado de Máquina e a partir de uma série de exemplos você pôde entender a sua importância em todos os setores da indústria.

Você deve ter entendido as diferenças entre **Inteligência Artificial** e o **Aprendizado de Máquina**, que trata de um tipo de Inteligência baseado em Dados. Isso cria um novo paradigma para o desenvolvimento de programas e a solução de problemas.

Você também aprendeu o que são e como diferenciar os Aprendizados de Máquina **Supervisionado**, **Não Supervisionado** e de **Reforço**, sendo característico de cada um a predição, a descrição ou a prescrição a partir de dados.

No restante deste curso trataremos unicamente modelos de Aprendizado Supervisionado e, particularmente, estaremos interessados em construir, analisar e explorar modelos de **Regressão (predição de valores)** e **Classificação (predição de classes ou categorias)** a partir de dados previamente rotulados, os **conjuntos de treinamento**.

# Para Saber Mais

* Acesse \* \_\_\_.**Google, Crash Course - ML Introduction**. em: <https://developers.google.com/machine-learning/crash-course/ml-intro>, e assista aos vídeos dos capítulos *Introduction to ML* (3min), *Framing* (15min) e *Descending into ML* (20min) para revisar boa parte dos conceitos que você aprendeu aqui em uma visão da Google.
* Que tal explorar alguns cases de aplicação do aprendizado de máquina dentro da sua área de interesse e compartilhar com os colegas? Acesse aqui \_\_\_.**Machine Learning and Data Science Applications in Industry** Disponível em: <https://github.com/ashishpatel26/Real-time-ML-Project> e explore alguns desses cases!
* Neste curso você vai entender e aplicar vários modelos de aprendizado aprendizado supervisionado. Que tal entender mais sobre esses modelos? Acesse, então \_\_\_.**What is Machine Learning?** Disponível em: <https://c3.ai/introduction-what-is-machine-learning/supervised-learning/> Acesso em: 02 de Novembro de 2021. Esse texto trás e explica exmplos de aprendizado supervisionado para classificação e regressão de dados.

# Referências

* Brown, Sata. **Machine learning, explained**. April, 2021. Disponível em: <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>. Acesso em: 04 de Março de 2022.
* Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, Rüdiger Wirth [**CRISP DM 1.0**](http://meusite.mackenzie.br/rogerio/TIC2021S1$/T1/crisp-dm.pdf) **Step-by-step data mining guide** Disponível em: <http://meusite.mackenzie.br/rogerio/TIC2021S1$/T1/crisp-dm.pdf> Acesso em: 29 de Agosto de 2021.
* Delua, Julianna **Supervised vs. Unsupervised Learning: What’s the Difference?** Disponível em: <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning> Acesso em: 02 de Novembro de 2021.
* Mwiti, Derrick **10 Real-Life Applications of Reinforcement Learning** Disponível em: <https://neptune.ai/blog/reinforcement-learning-applications> Acesso em: 02 de Novembro de 2021.
* Kotu, Vijay; Deshpande, Balachandre **Data Science: concepts and practice**. 2nd ed. Cambridge, [England]: Morgan Kaufmann, c2019. E-book (570 p.) ISBN 9780128147627 (electronic bk.). Disponível em: <http://pergamum.mackenzie.br:8080/pergamumweb/vinculos/00003c/00003cef.jpg>.
* Alpaydin, E. **Machine Learning** (The MIT Press Essential Knowledge). The MIT Press. 2019.
* Kelleher,John D.; Tierney, Brendan. **Data Science** (The MIT Press Essential Knowledge). The MIT Press. 2018.
* \_\_\_.**What is Machine Learning?** Disponível em: <https://c3.ai/introduction-what-is-machine-learning/supervised-learning/> Acesso em: 02 de Novembro de 2021.
* \_\_\_.**Machine Learning and Data Science Applications in Industry** Disponível em: <https://github.com/ashishpatel26/Real-time-ML-Project> Acesso em: 02 de Novembro de 2021.
* \_\_\_.**Google, Crash Course - ML Introduction**. Disponível em: <https://developers.google.com/machine-learning/crash-course/ml-intro>. Acesso em: 03 de Março de 2022.