

MINICURSO

INTRODUÇÃO AO APRENDIZADO DE MÁQUINAS SUPERVISIONADO

UMA APLICAÇÃO EM CONSUMO ENERGÉTICO

Professora Angélica Lourenço Oliveira

Doutora em Matemática Aplicada - Inteligência Computacional

Semana da Engenharia Elétrica - UEMG Ituiutaba

CONTEÚDO

- INTELIGÊNCIA ARTIFICIAL
- CIÊNCIA DE DADOS
- APRENDIZADO DE MÁQUINAS
- TÉCNICAS DE APRENDIZADO
- MÉTRICAS DE AVALIAÇÃO
- APLICAÇÃO PRÁTICA

Acesse:

https://github.com/angelicalo/eletrica_uemg

<https://shre.ink/cursouemg>

INTELIGÊNCIA ARTIFICIAL



Inteligência Humana é uma combinação de habilidades cognitivas, sociais e emocionais que permitem ao ser humano interagir com o mundo ao redor.

Podemos reconhecer padrões com base no aprendizado ao longo da vida.

INTELIGÊNCIA ARTIFICIAL



Inteligência não-humana embora os seres humanos sejam o foco primordial da pesquisa em inteligência, a inteligência exibida por outros animais também é alvo de investigação.

As plantas também podem ser consideradas inteligentes: se reconhecerem mutuamente, monitorarem o ambiente e ajustarem sua morfologia, fisiologia e fenótipo de acordo com as circunstâncias.

INTELIGÊNCIA ARTIFICIAL



A **Inteligência Artificial - IA** (ou AI, de artificial intelligence), é um conjunto de algoritmos e técnicas que permitem que um sistema computacional aprenda a realizar uma determinada tarefa — indo além da programação de ordens específicas para tomar decisões de forma autônoma, baseadas em padrões de bancos de dados.

INTELIGÊNCIA ARTIFICIAL



- Chatbots
- Eficiência em Gestão
- Assistente Pessoal
- Segurança da Informação
- Predições
- Vendas e Marketing
- Educação e Ensino

INTELIGÊNCIA ARTIFICIAL



Implementar a inteligência artificial nas empresas pode *potencializar os resultados*, deixar os processos mais ágeis e ajudar a *reduzir gastos operacionais*. Além das vantagens, a *experiência do usuário é aprimorada*, de maneira que os dados são muito mais acessíveis e corretos. Os relatórios gerados são *mais próximos da realidade*, ajudando a criar *insights* que melhorem o processo final.

INTELIGÊNCIA ARTIFICIAL



Reconhecimento de
padrões:

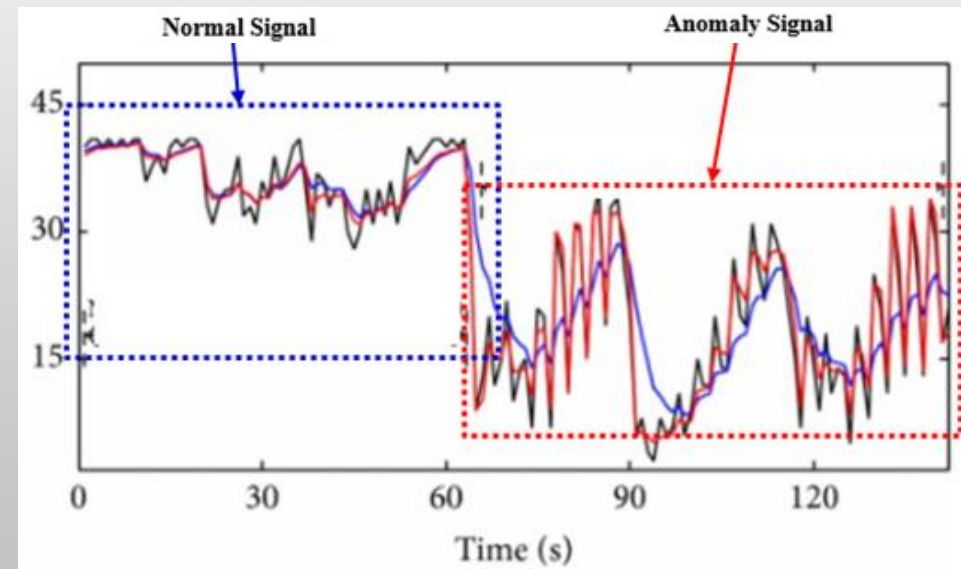
INTELIGÊNCIA ARTIFICIAL



Reconhecimento de
padrões:

Existe um padrão?

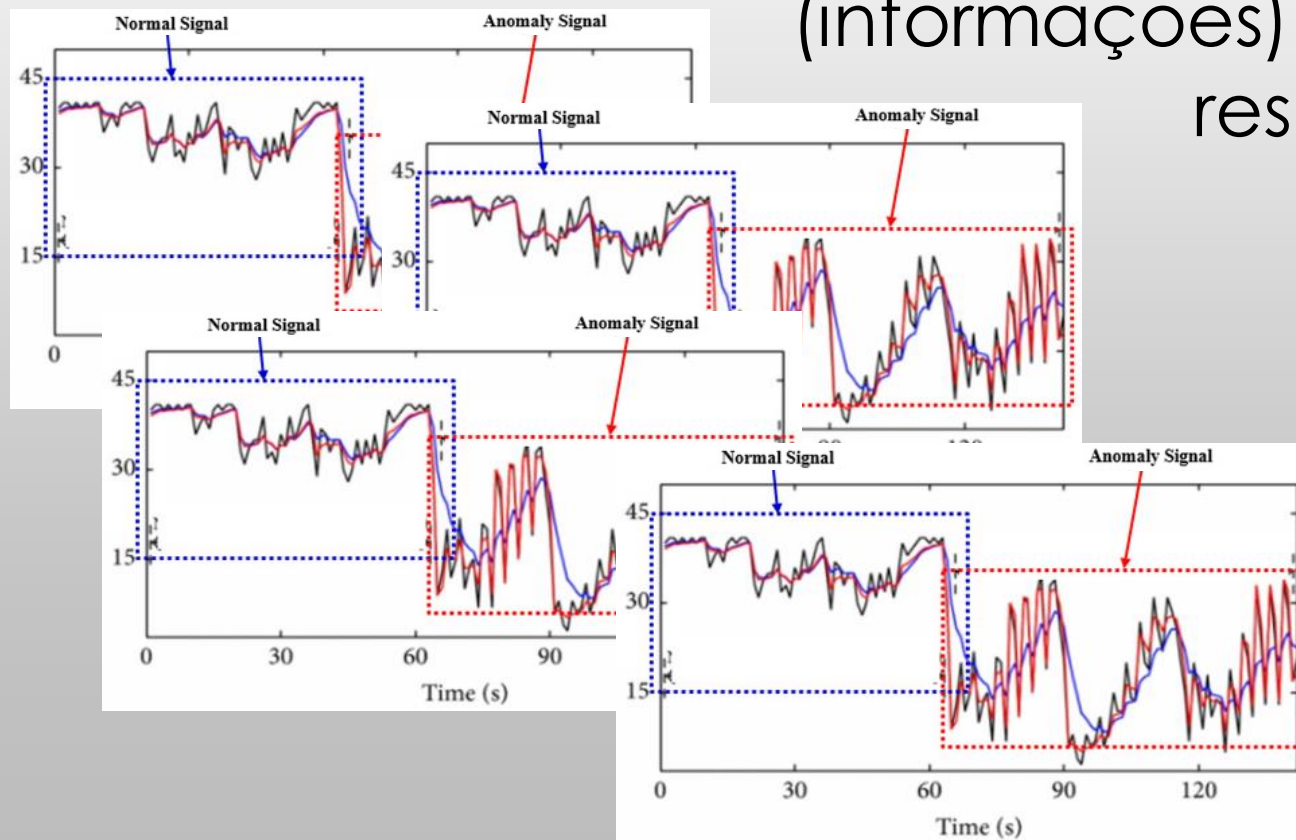
INTELIGÊNCIA ARTIFICIAL


















INTELIGÊNCIA ARTIFICIAL



Mais experiência
(informações) melhor
resultado?



The companies that know most about you

#	Company	% of personal data collected	Email	Name	Age	Gender	Sexual	Marital	Race	Religion	Live Loc	Home	Emplo	Job Title	Pet/Animal	Mobile	Landline	Type of Device	Hobbies	Interests	Height	Weight	Next of Kin	Mother's Name	Current Employer	Past Employers	Bank / Finance	Salary	Social Media	Social Media	Social Media	Count	Allergies	Health	
1	 Facebook	70.59%	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•		•	•			•	•	•			
2	 Instagram	58.82%	•	•	•	•					•	•	•	•	•	•		•	•	•	•	•		•	•			•	•	•					
3	 Tinder	55.88%	•	•	•	•	•	•			•		•	•	•	•		•	•	•	•					•		•	•	•					
4	 Grindr	52.94%	•	•	•	•	•	•			•		•	•	•	•		•	•	•	•						•	•	•						
5	 Uber	52.94%	•	•	•	•					•	•	•	•		•	•	•	•	•					•	•		•	•	•					
6	 Strava	41.18%	•	•	•	•					•	•				•		•			•	•				•		•	•	•					
7	 Tesco	38.24%	•	•	•	•		•				•				•	•	•								•		•	•	•					
8	 Spotify	35.29%	•	•	•							•				•	•	•		•						•		•	•	•					
9	 MyFitnessPal	35.29%	•	•	•							•				•	•	•			•	•				•						•			
10	 Jet2	35.29%	•	•	•	•		•	•			•				•	•	•								•					•				
11	 Credit Karma	32.35%	•	•	•								•	•		•		•			•	•				•	•								
12	 Lidl Plus	32.35%	•	•	•						•	•			•	•	•	•								•						•			
13	 Netflix	26.47%	•	•	•							•				•	•	•		•						•									
14	 Nike	26.47%	•	•	•	•						•						•			•	•				•									
15	 Asos	26.47%	•	•	•							•				•		•			•	•				•									

INTELIGÊNCIA ARTIFICIAL



INTELIGÊNCIA ARTIFICIAL

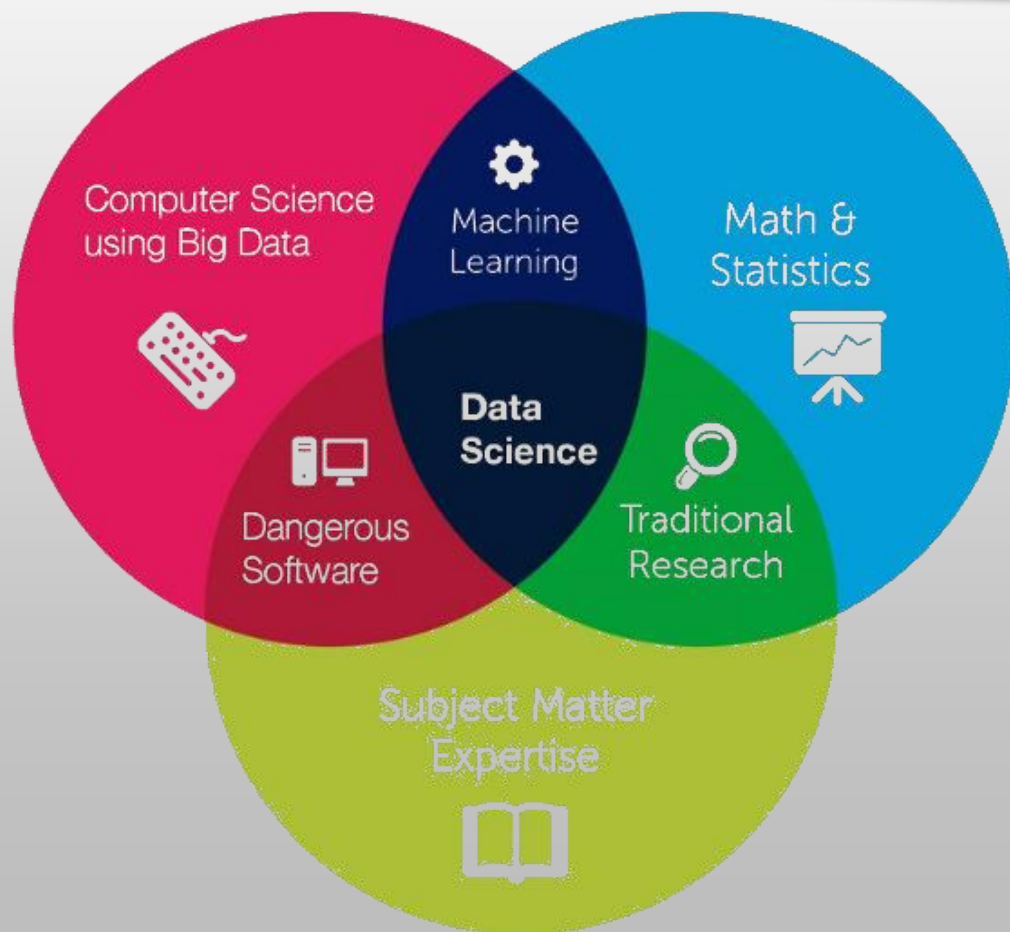


Reconhecimento facial



Dados são considerados o novo Petróleo.

CIÊNCIA DE DADOS

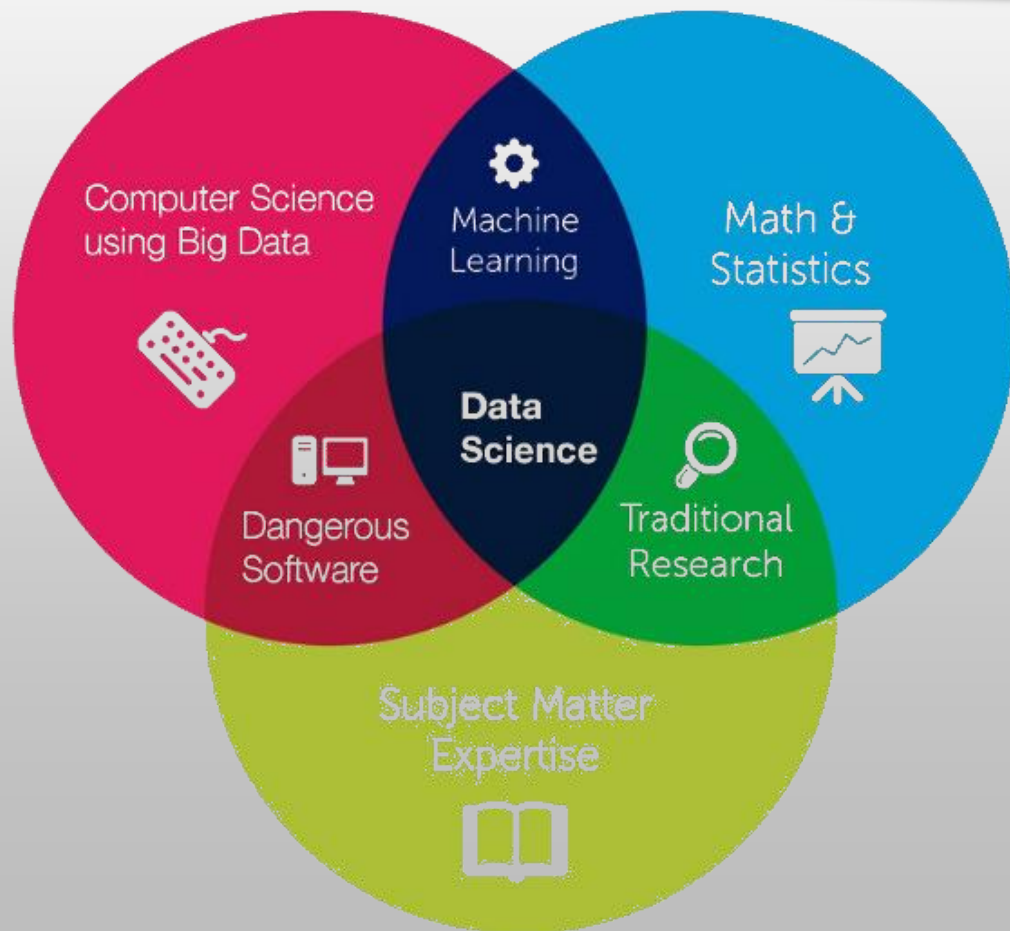


Dados são considerados o novo Petróleo.

[Consumo Mensal de Energia Elétrica por Classe \(regiões e subsistemas\) \(epe.gov.br\)](https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/consumo-de-energia-eletrica)

<https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/consumo-de-energia-eletrica>

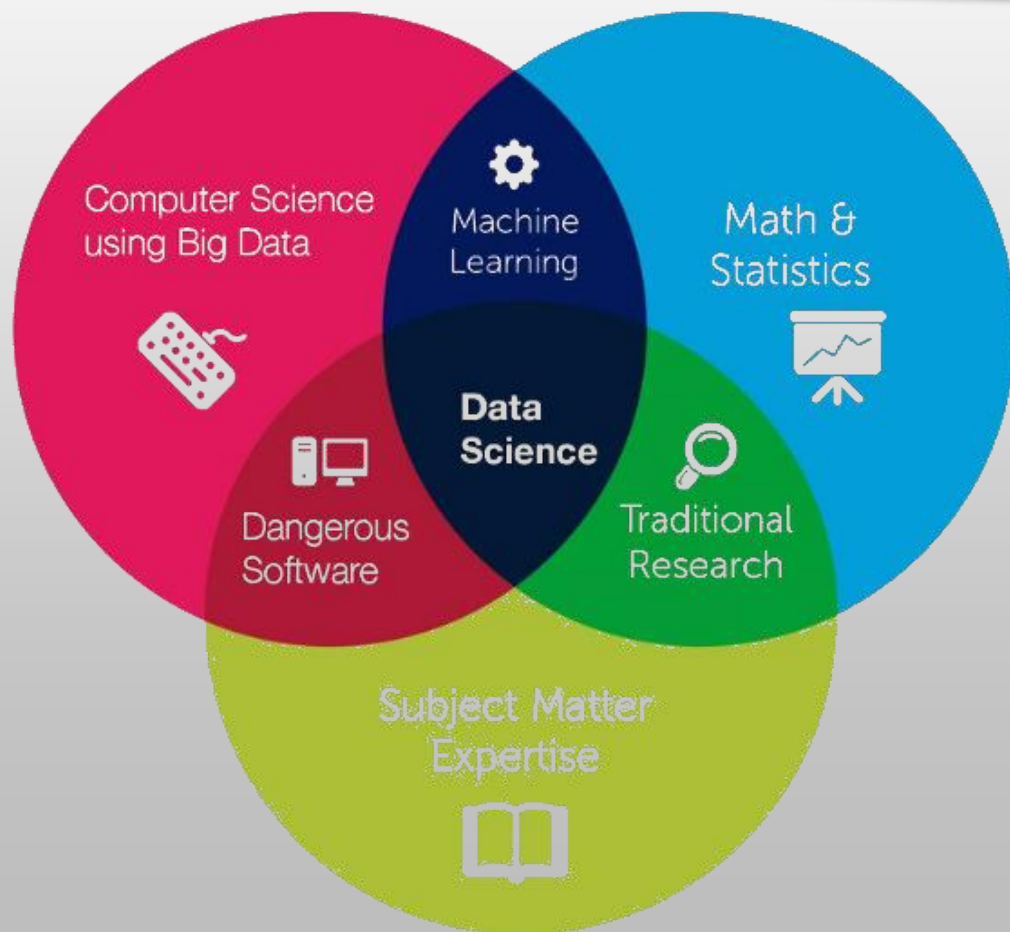
CIÊNCIA DE DADOS



Dados são considerados o novo Petróleo.

O **profissional/empresa** que consegue extrair, ler e manipular bem os dados, tem "Petróleo" em suas mãos!

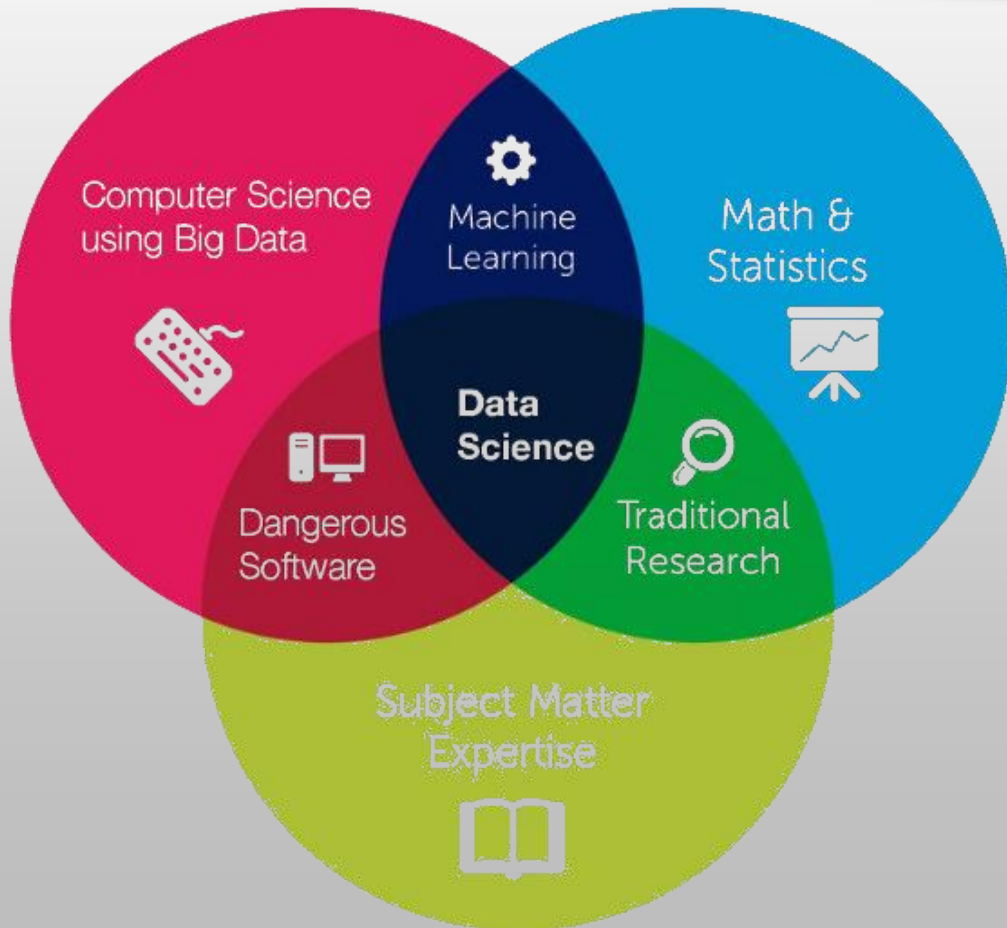
CIÊNCIA DE DADOS



A ciência de dados é um processo que aplica o método científico no processo de análise de dados a fim de alcançar um objetivo.

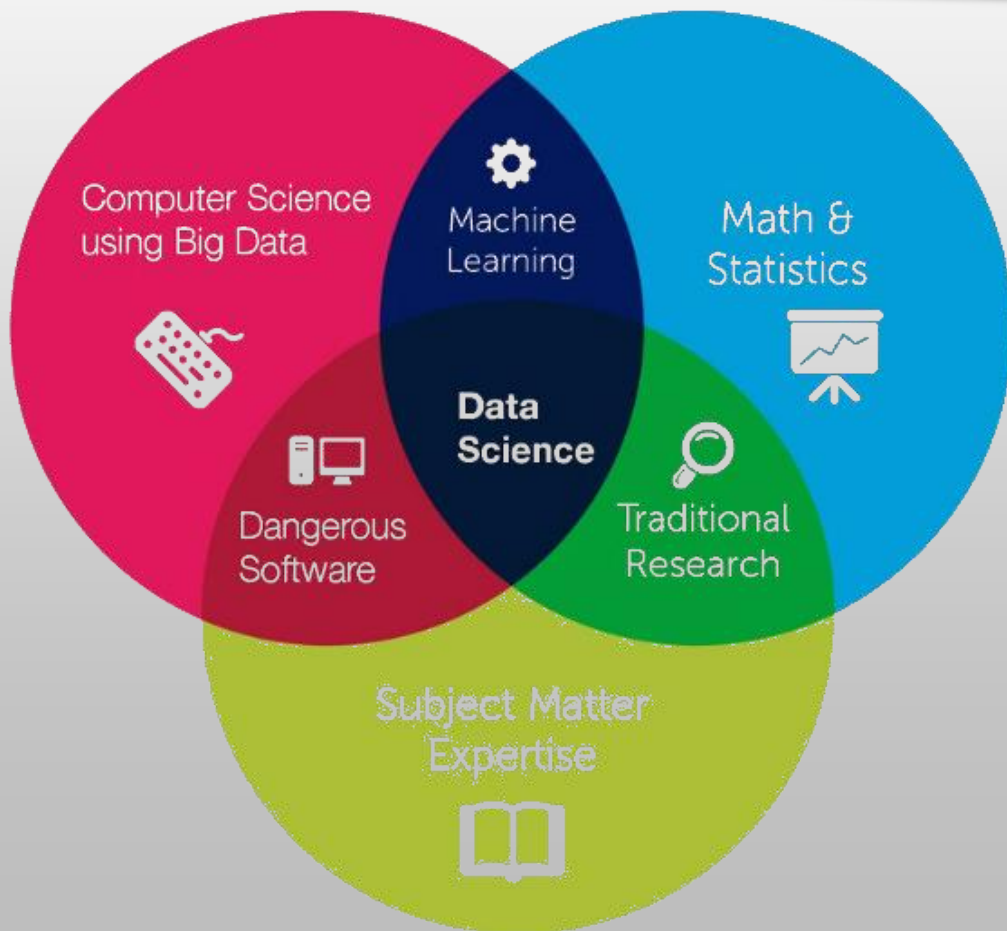
O profissional responsável por fazer este estudo é o Cientista de Dados.

CIÊNCIA DE DADOS



O que eu preciso dominar para atuar como Cientista de Dados?

CIÊNCIA DE DADOS

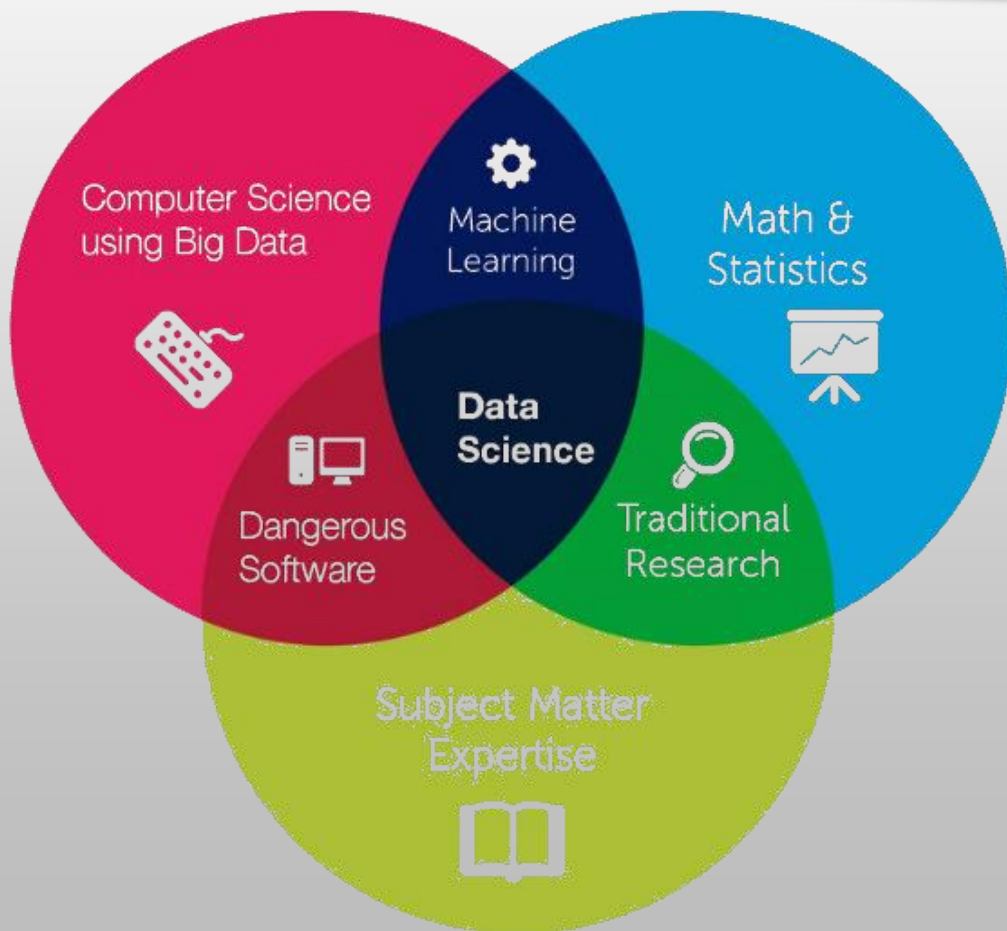


O que eu preciso dominar para atuar como Cientista de Dados?

Tudo!



CIÊNCIA DE DADOS

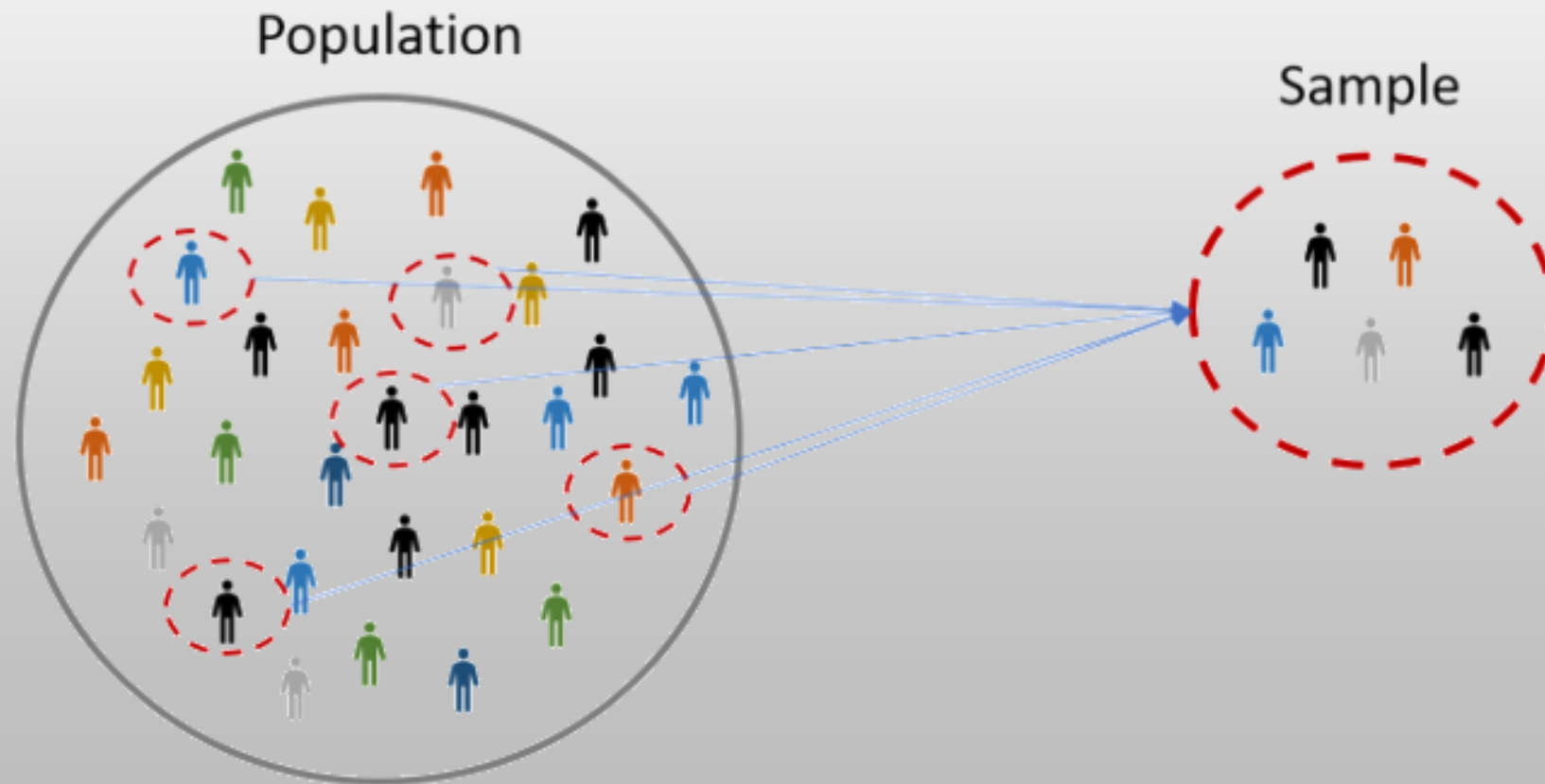


O que eu preciso dominar para atuar como Cientista de Dados?

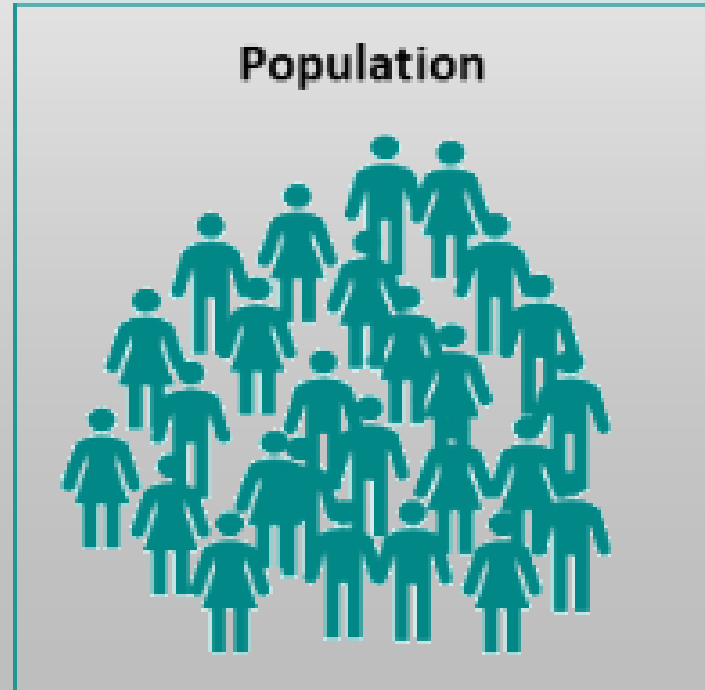
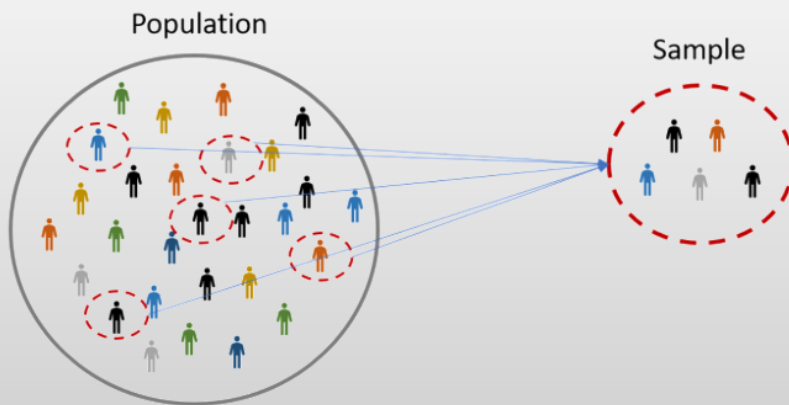
~~Tudo!~~



CIÊNCIA DE DADOS

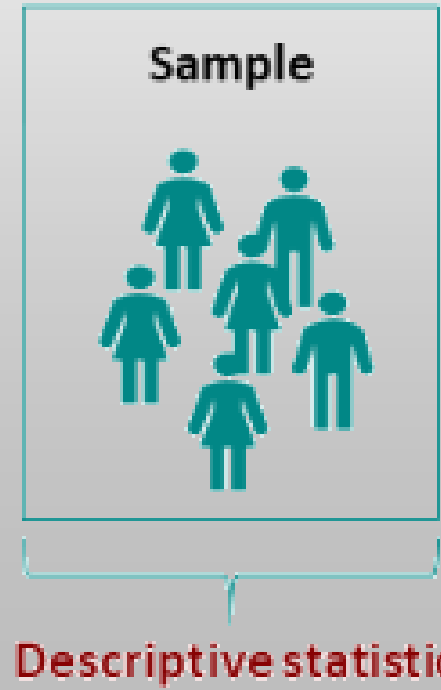


CIÊNCIA DE DADOS



Sampling

A teal arrow pointing from the Population box to the Sample box, labeled 'Sampling'.



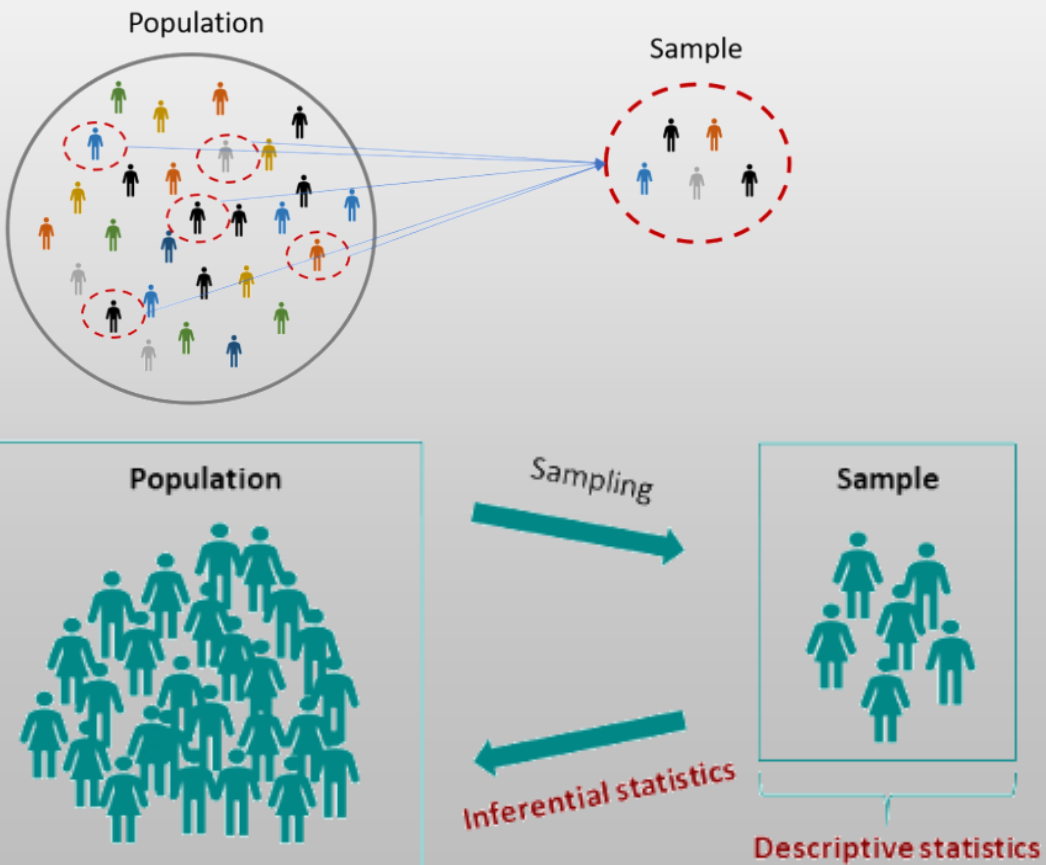
Inferential statistics

A teal arrow pointing from the Sample box back to the Population box, labeled 'Inferential statistics'.

Descriptive statistics

A bracket under the Sample box, labeled 'Descriptive statistics'.

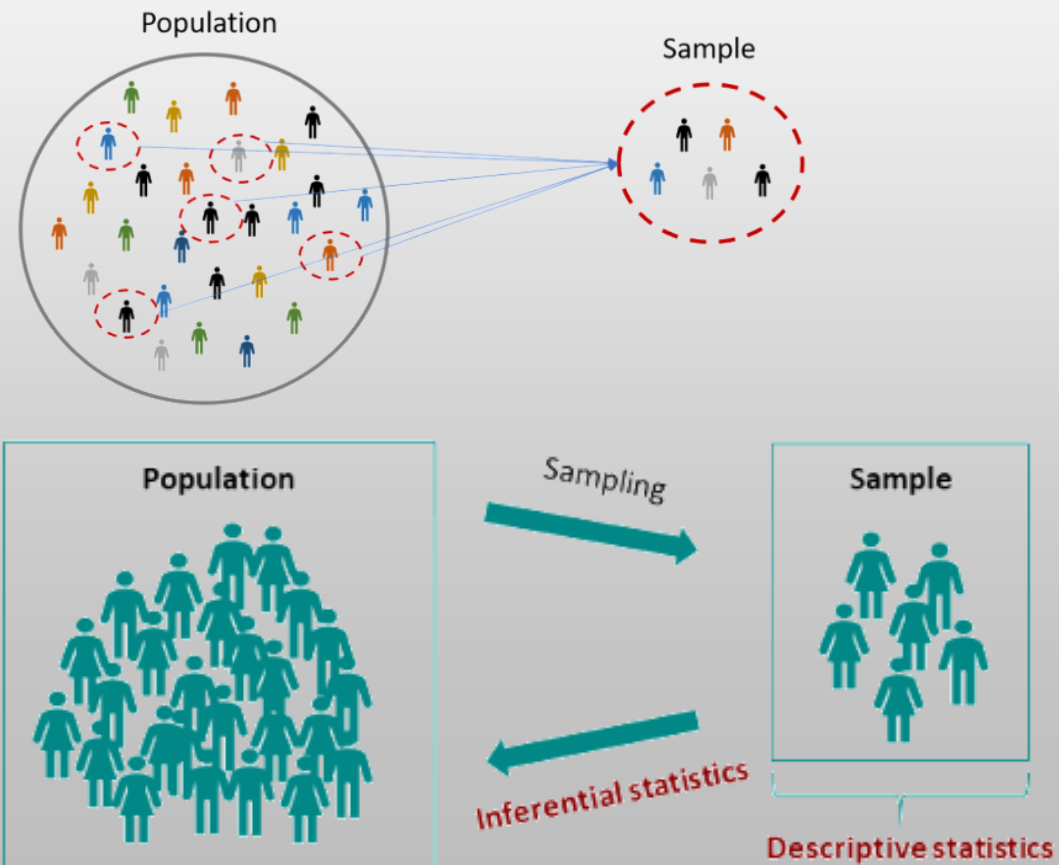
CIÊNCIA DE DADOS



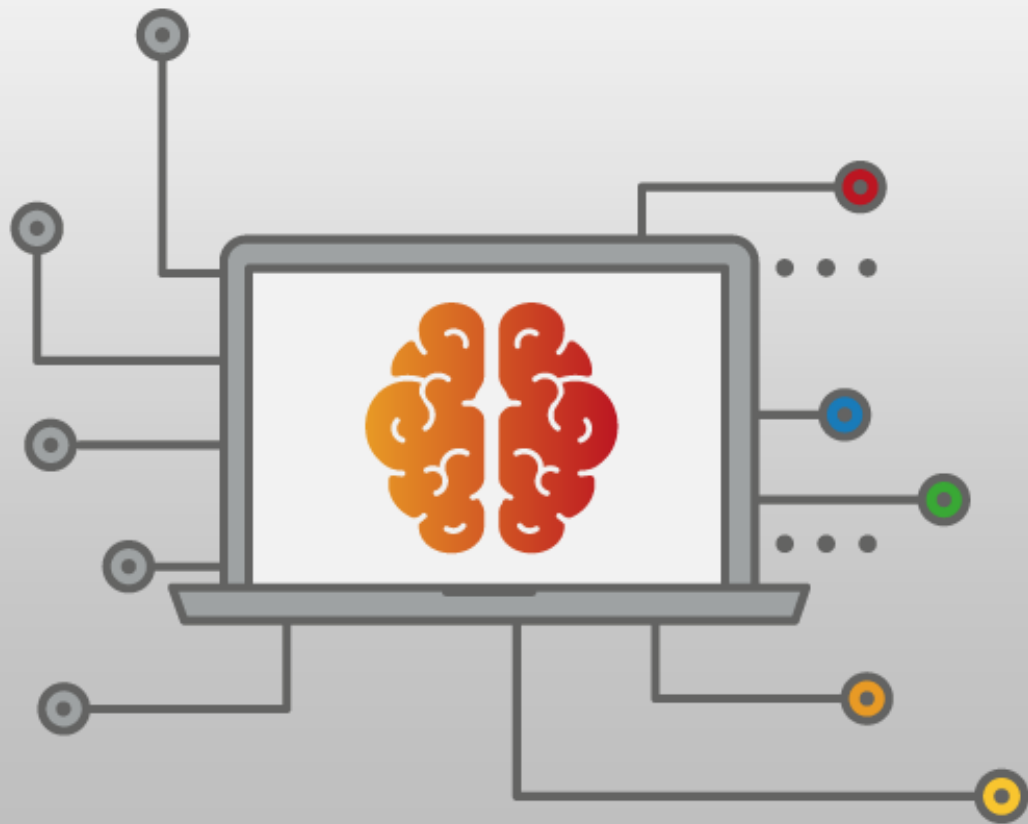
Amostragem aleatória simples (python)

CIÊNCIA DE DADOS

Quantidade
vs
Qualidade



APRENDIZADO DE MÁQUINAS



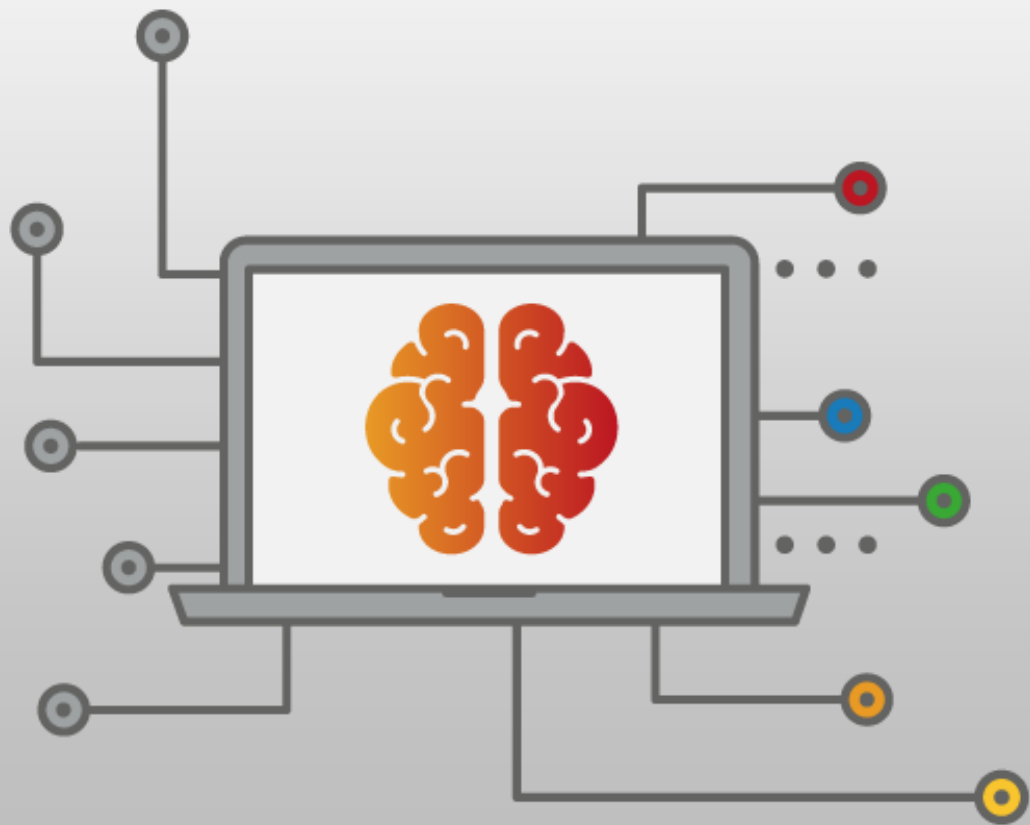
Quantidade
vs
Qualidade

APRENDIZADO DE MÁQUINAS

Quantidade
vs
Qualidade

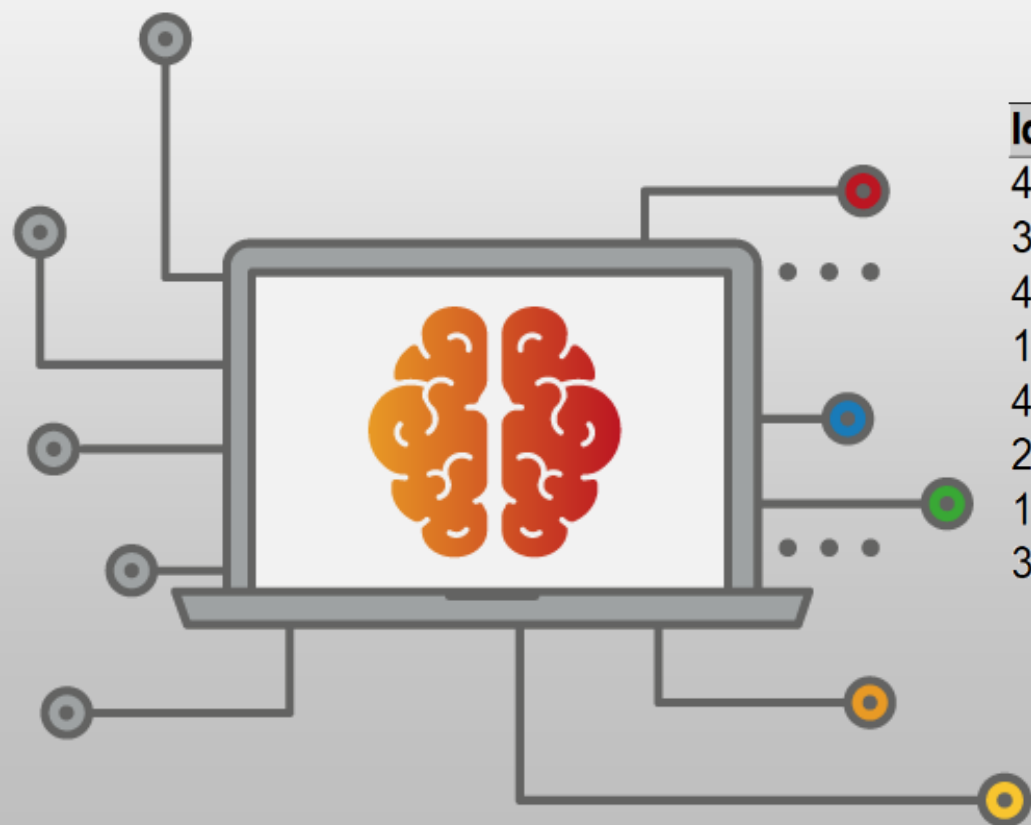


APRENDIZADO DE MÁQUINAS



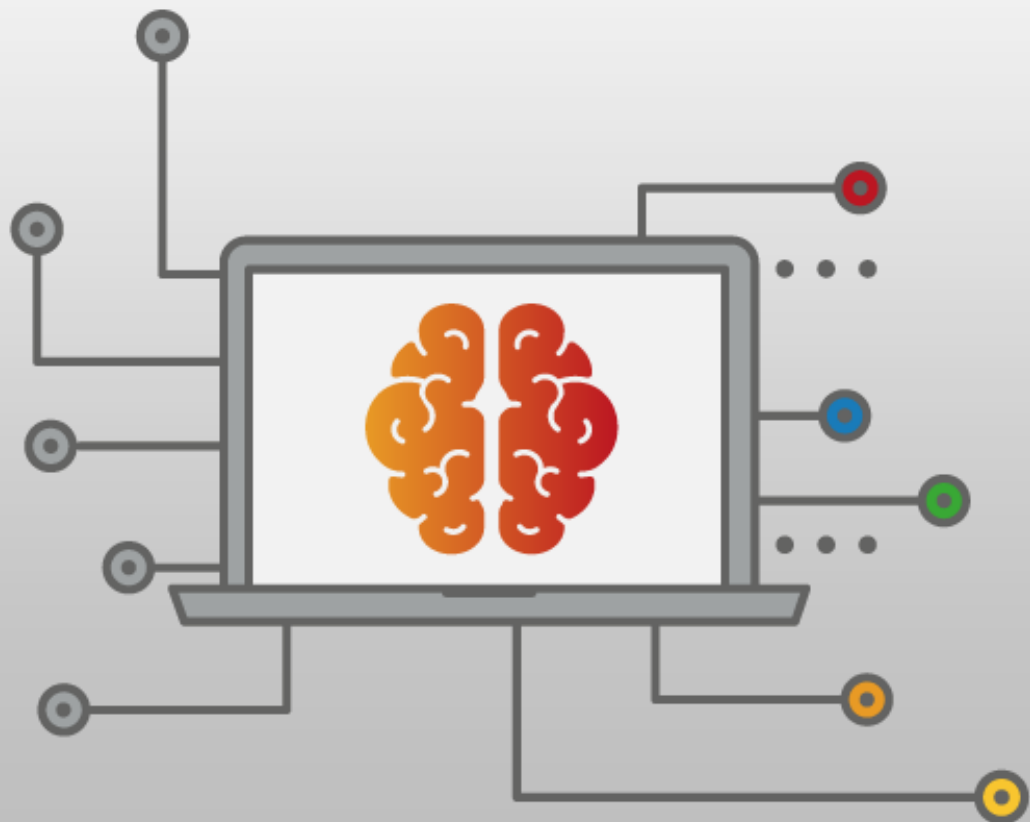
Tratamento dos
dados
=
Qualidade

APRENDIZADO DE MÁQUINAS



Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

APRENDIZADO DE MÁQUINAS

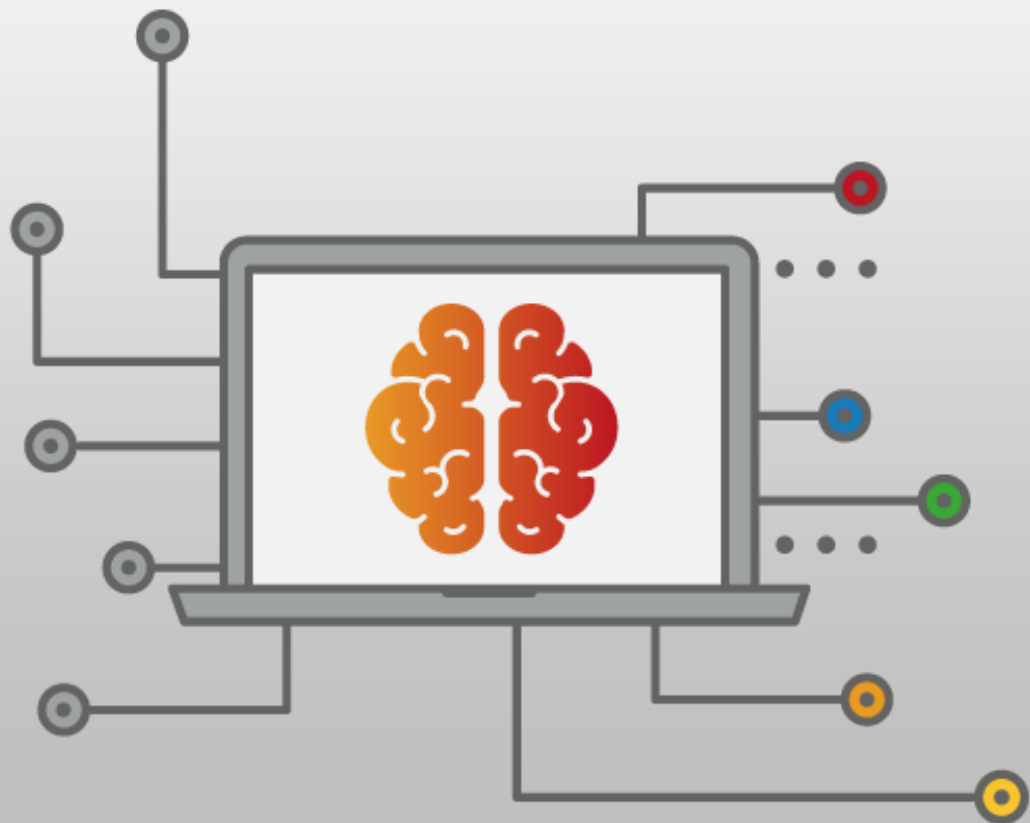


Imagine escrever um programa de computador para:

- Diagnosticar paciente a partir de seus sintomas
- Responder questões sobre vendas
- Direcionar campanhas de descontos para clientes
- Identificar equipamentos com chances de falha durante a produção

A forma com que o computador será ensinado a fazer essas tarefas é o que chamamos de aprendizado de máquinas.

APRENDIZADO DE MÁQUINAS

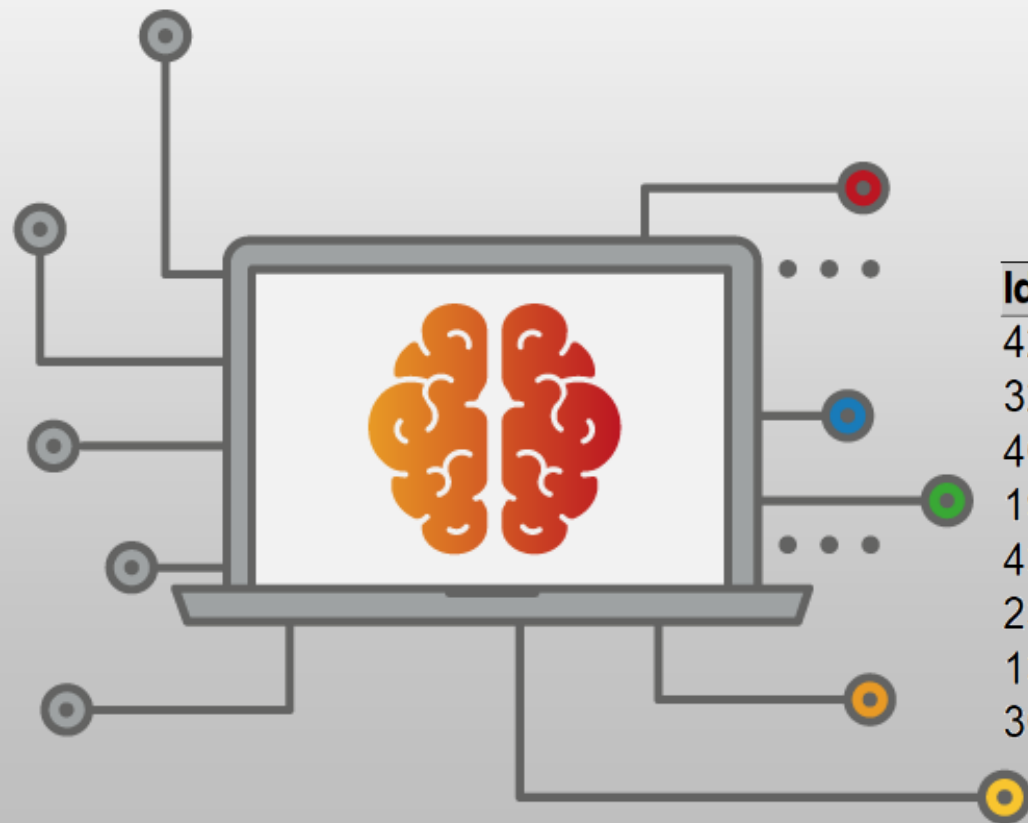


Aprendizado de máquinas

Técnicas capazes de criar, a partir de experiência passada. Uma hipótese (função) capaz de resolver um problema.

Um programa aprende a partir da experiência E , em relação a uma classe de tarefas T , com medida de desempenho P , se seu desempenho em T , medido por P , melhora com E .

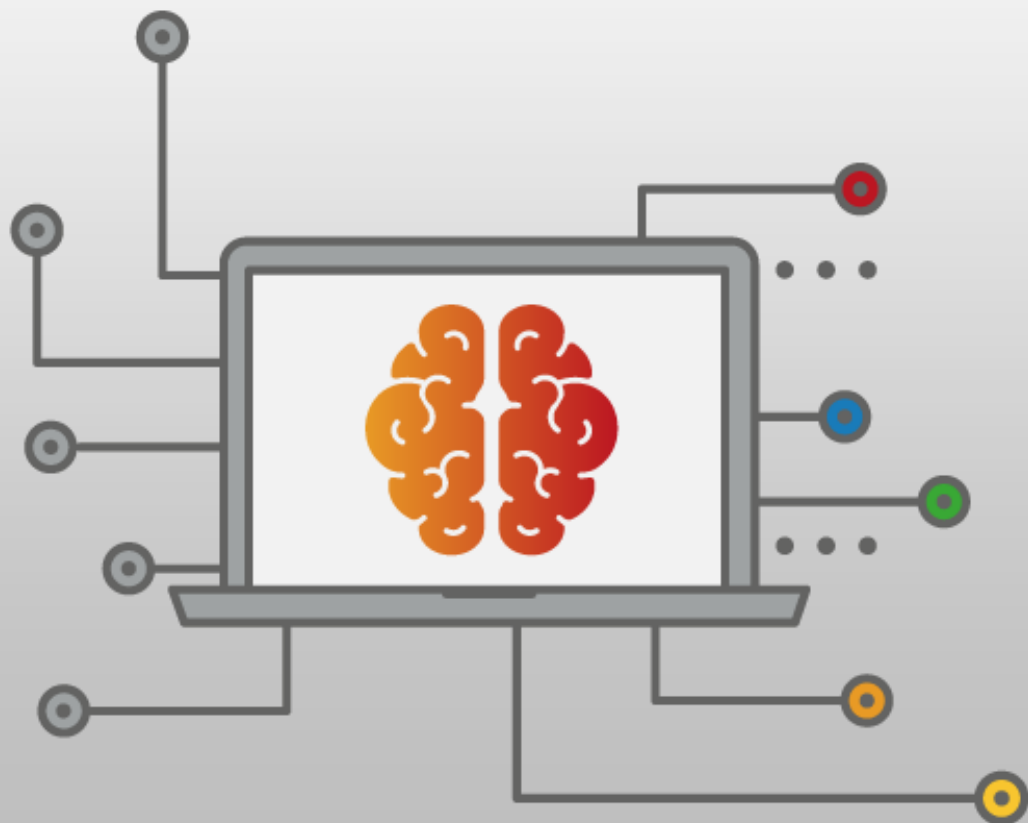
APRENDIZADO DE MÁQUINAS



Saída

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

APRENDIZADO DE MÁQUINAS



Problema - Reconhecer escrita manual

Tarefa T:

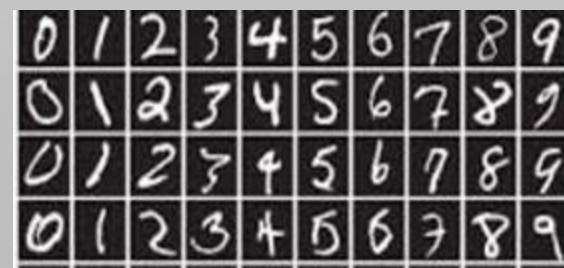
Classificar dígitos corretamente identificados

Medida de desempenho P:

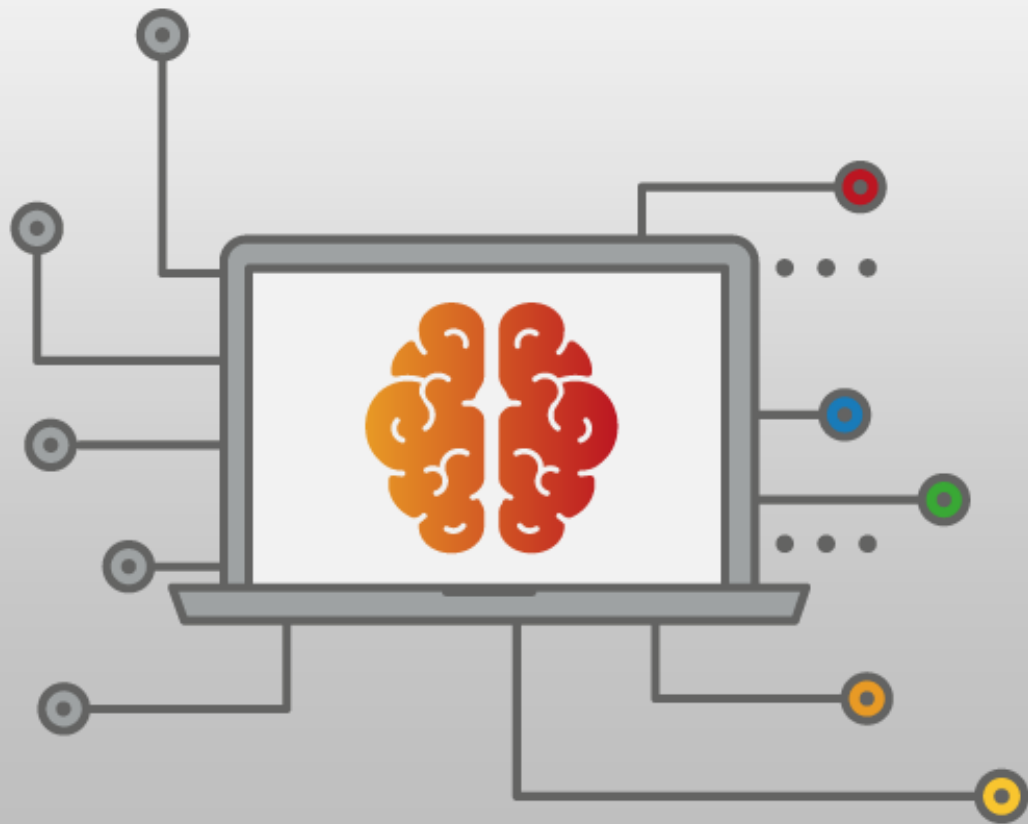
Porcentagem de dígitos corretamente identificados

Experiência E:

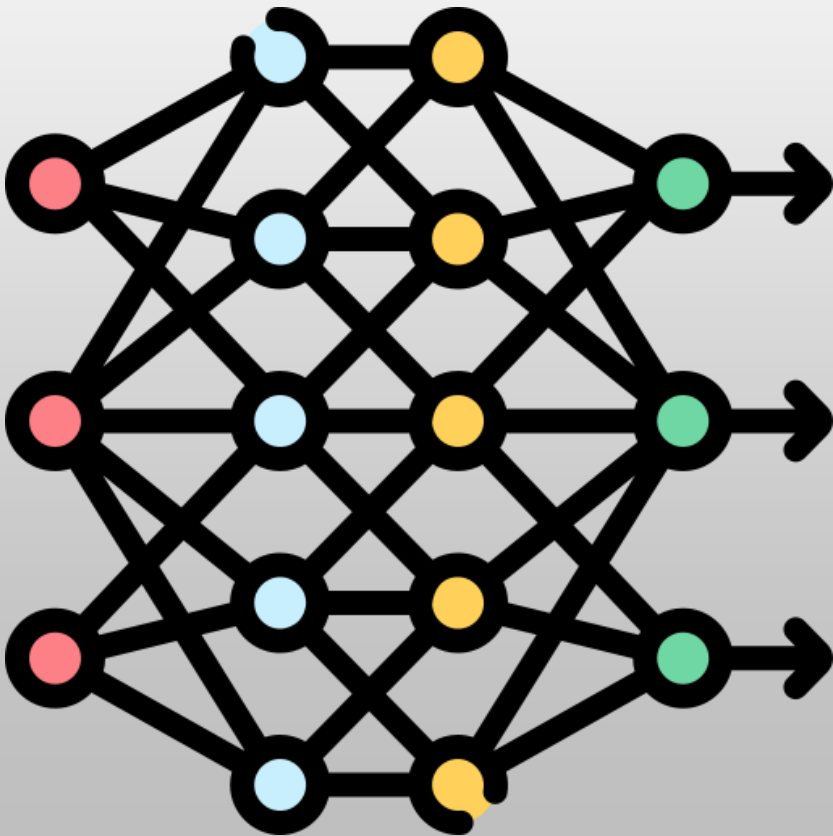
Imagens de dígitos manuscritos por pessoas diferentes



APRENDIZADO DE MÁQUINAS



TÉCNICAS DE APRENDIZADO

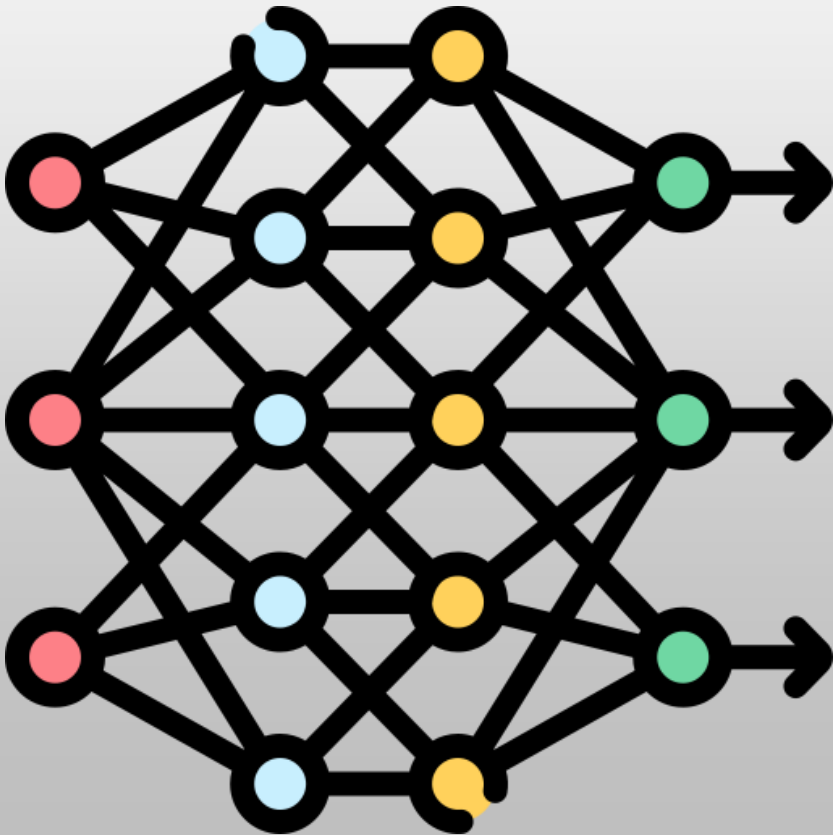


Principais tipos de aprendizado

Supervisionado

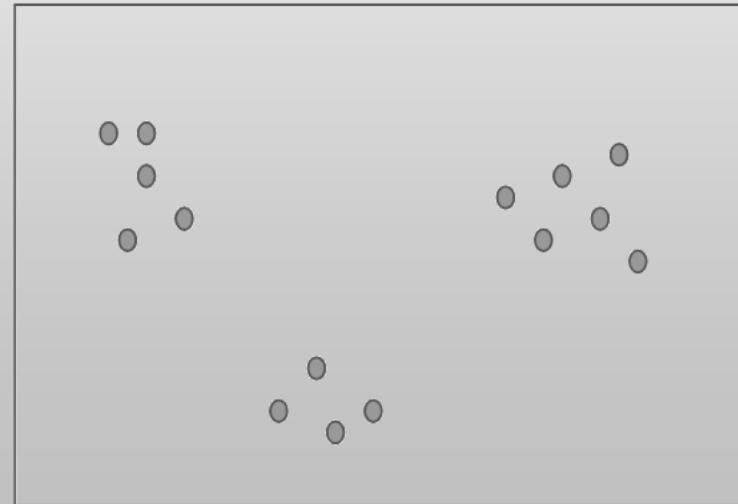
Não-supervisionado

TÉCNICAS DE APRENDIZADO

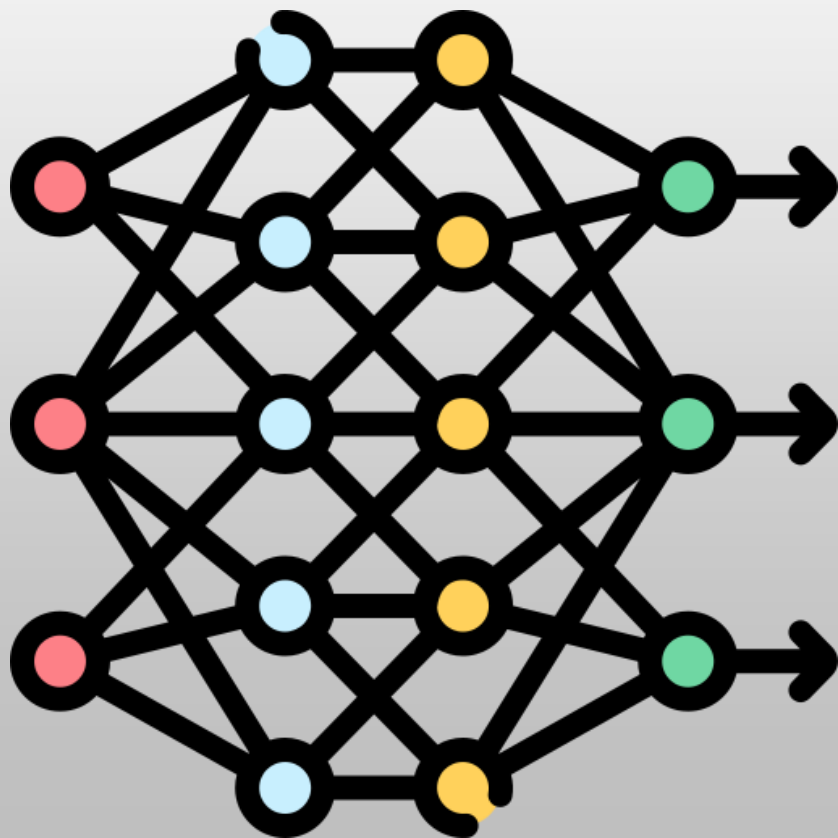


Não supervisionado

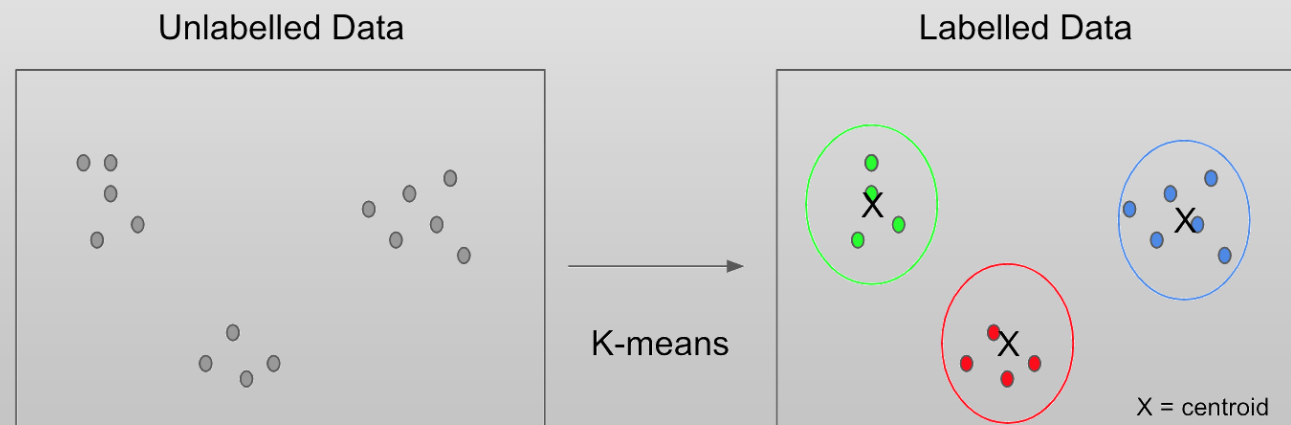
Unlabelled Data



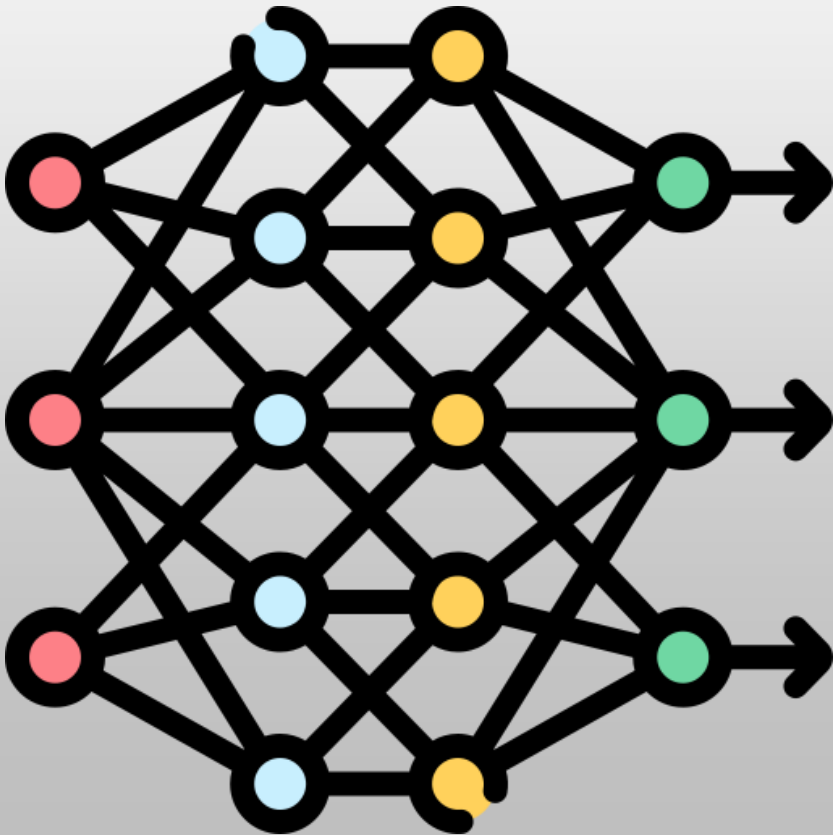
TÉCNICAS DE APRENDIZADO



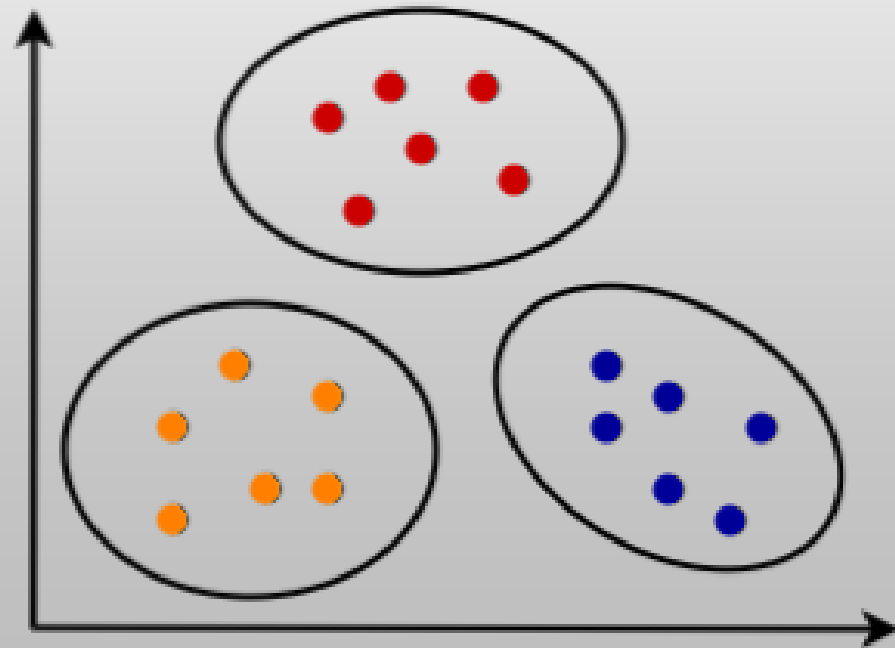
Não supervisionado



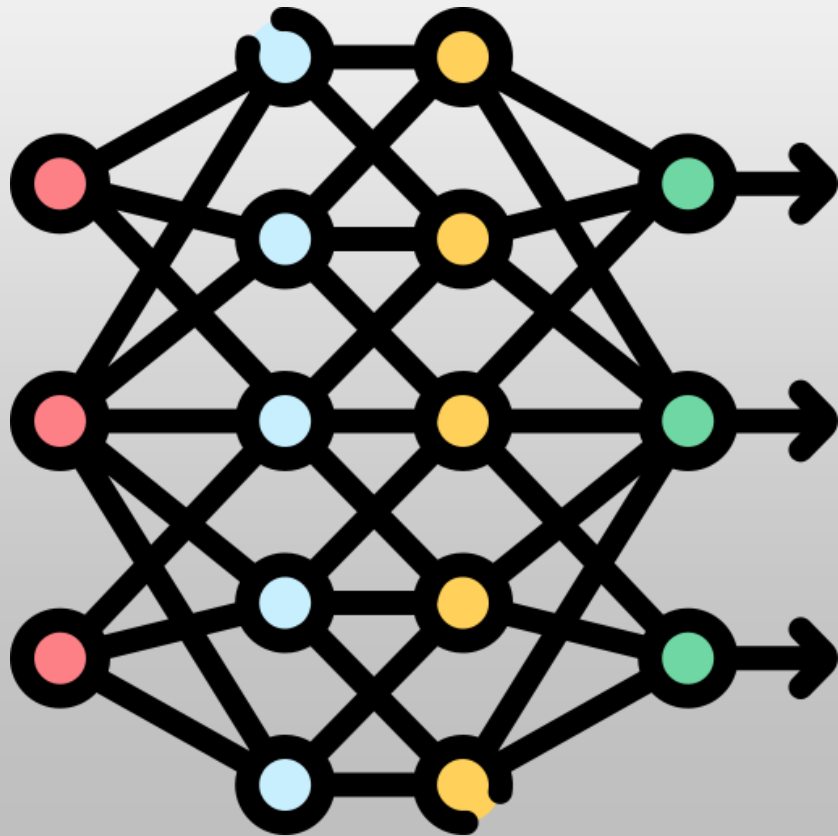
TÉCNICAS DE APRENDIZADO



Supervisionado

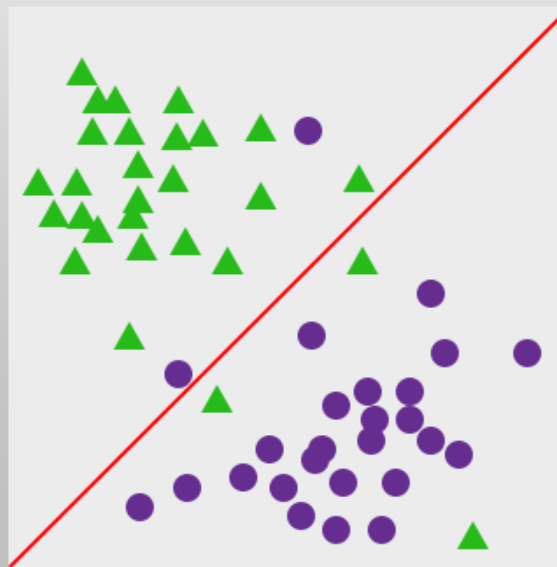


TÉCNICAS DE APRENDIZADO



Supervisionado

CLASSIFICAÇÃO



REGRESSÃO



MÉTRICAS DE AVALIAÇÃO



Medidas de avaliação são utilizadas para quantificar o desempenho de modelos de aprendizado de máquina em tarefas de classificação e regressão.

MÉTRICAS DE AVALIAÇÃO



Algumas medidas comumente usadas são:

Classificação:

Acurácia: Mede a proporção de exemplos classificados corretamente em relação ao total de exemplos.

Precisão: Mede a proporção de exemplos classificados corretamente como positivos em relação ao total de exemplos classificados como positivos.

Recall (Revocação): Mede a proporção de exemplos positivos classificados corretamente em relação ao total de exemplos positivos.

F1-Score: Combina a precisão e o recall em uma única medida, levando em consideração tanto os verdadeiros positivos quanto os falsos positivos.

Matriz de Confusão: Tabela que mostra a contagem de exemplos classificados corretamente e erroneamente para cada classe.

MÉTRICAS DE AVALIAÇÃO



Algumas medidas comumente usadas são:
Classificação:

$$\text{Acurácia} = \frac{\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos}}{\text{Total de Exemplos}}$$

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos}}$$

$$\text{Recall} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Negativos}}$$

$$\text{F1-Score} = \frac{2 \times \text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}}$$

MÉTRICAS DE AVALIAÇÃO



MATRIZ DE CONFUSÃO

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

MÉTRICAS DE AVALIAÇÃO



Algumas medidas comumente usadas são:

Regressão:

Erro Médio Absoluto (MAE): Calcula a média das diferenças absolutas entre as previsões e os valores reais.

Erro Quadrático Médio (MSE): Calcula a média das diferenças quadráticas entre as previsões e os valores reais.

Raiz do Erro Quadrático Médio (RMSE): Calcula a raiz quadrada do MSE, fornecendo uma medida na mesma escala que a variável alvo.

Coeficiente de Determinação (R^2): Mede a proporção da variância da variável alvo que é explicada pelo modelo. Um valor próximo de 1 indica um bom ajuste.