

## 1 Introdução aos Módulos de Machine Learning

# Introdução aos Módulos de *Machine Learning*

## Inteligência Artificial

A inteligência artificial (IA) é um conceito amplo, porém comumente definido como mecanismos computacionais que se baseiam no comportamento humano para resolver problemas. Em outras palavras, a tecnologia faz o computador pensar como os seres humanos para execução de atividades.

Nós seres humanos estamos constantemente analisando dados e informações em qualquer lugar que vamos, encontramos padrões como também tendências nestas análises que nos direcionam para conclusões ou decisões. De certa maneira, a IA segue este mesmo princípio.

Assim como nós, seres humanos, à medida que executamos com maior frequência uma tarefa, mais habilidosos ficamos e passamos a oferecer maior agilidade nas nossas decisões como também maior assertividade. Isto é o reflexo da nossa capacidade de aprender. A repetição é uma grande aliada no processo de aprendizagem que funciona como um treinamento para nós.

Algo bastante semelhante ocorre nos sistemas de IA, como por exemplo, silos de dados em diferentes áreas da empresa, dados centralizados em data lakes, dados em bancos de dados ou dados localizados em qualquer tipo ou plataforma são utilizados como ativos ou recursos computacionais que permitem que algoritmos de sistemas de IA sejam treinados.

Não existe uma única maneira para realização deste treinamento. A realização do treinamento depende da aplicação, organização e pessoas envolvidas, como também novas tecnologias criadas que permitem a evolução ou reciclagem das técnicas existentes. Neste momento, o termo aprendizado de máquina ou *machine learning* surge e passa a fazer sentido.

## *Machine Learning*

O aprendizado de máquina ou *machine learning* é um sistema que pode modificar seu comportamento de forma autônoma com base na sua própria experiência. Esta experiência é o treinamento abordado anteriormente. A interferência humana é mínima, quando se desenvolve um sistema de aprendizado de máquina quebramos a estrutura convencional de programação. Na estrutura convencional é criado um conjunto de regras para gerar uma resposta a partir do processamento dos dados introduzidos.

Já em sistemas de aprendizado de máquina, estes são compostos por algoritmos, ou seja, não existem regras pré-estabelecidas, estas são geradas com base no conhecimento extraído a partir dos dados. Estes algoritmos fazem a análise dos dados e geram respostas ou resultados,

em que eles são utilizados para que o sistema de aprendizado de máquina crie suas próprias perguntas e regras.

Sobre as perguntas e regras geradas, estas visam melhorar o desempenho das tarefas de análise de dados com base na identificação de padrões ou comportamentos nos dados. Se pensarmos, por exemplo, em um sistema de busca, em que o usuário procura por esportes.

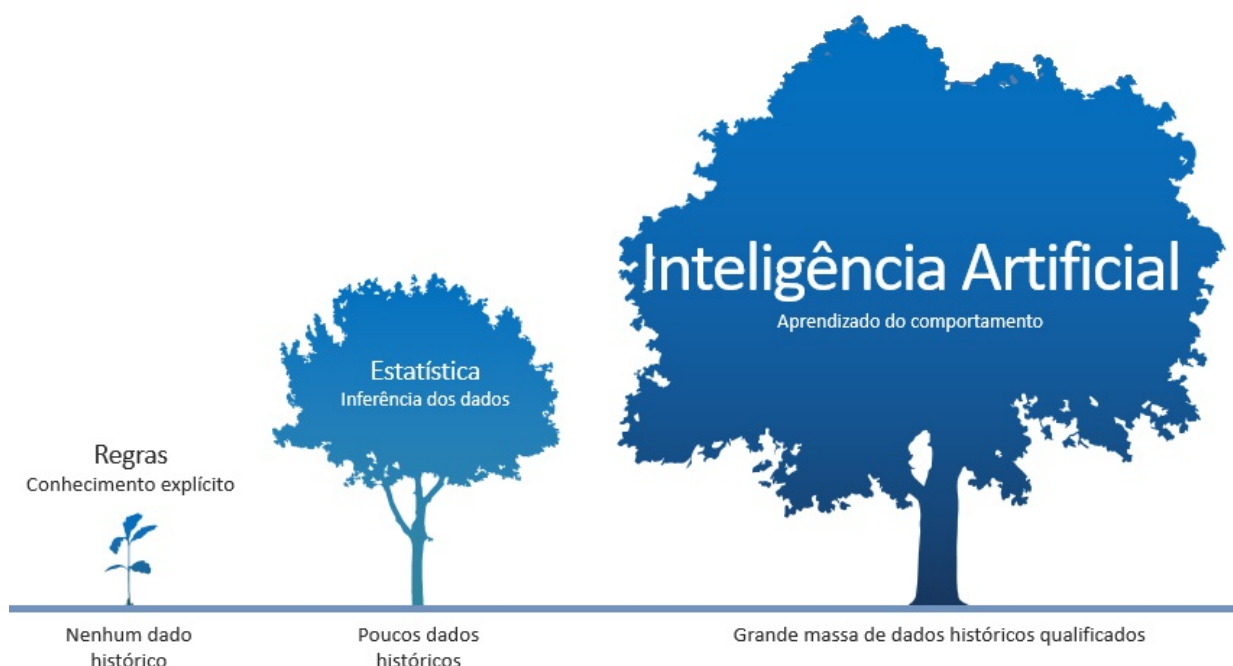
O sistema precisa analisar diversos parâmetros para entregar os resultados, como por exemplo futebol, basquete, vôlei etc. Dentre os parâmetros analisados, o histórico de pesquisa também é levado em consideração, se por exemplo, antes da pesquisa por esporte o usuário tivesse procurado por seleção brasileira, pentacampeã, Ronaldinho, Pelé e Neymar, o resultado para a busca mais provável para a pesquisa esporte seria futebol.

Apesar de simplista, o exemplo ilustra características importantes do *machine learning*. A quantidade de dados é um fator extremamente relevante e significativo para as análises de padrões e comportamento dos dados para os sistemas de *machine learning*.

Se considerarmos um cenário com poucos dados históricos de clientes de uma empresa do setor financeiro, a eficiência de sistemas de *machine learning* pode ser comprometida para a identificação dos padrões e comportamentos destes clientes e, neste cenário, as regras são criadas com base no conhecimento explícito sobre estes clientes que os funcionários possuem.

Ao passo que vamos evoluindo e conseguindo aumentar a captação de dados sobre estes clientes e a massa histórica aumenta, inferências sobre os dados podem ser feitas para produzir hipóteses utilizando as estatísticas. Muitas vezes, apenas utilizando estatísticas sem a necessidade de modelagem de dados já nos permite entender as vertentes de negócio e conduzir uma conclusão sobre os dados.

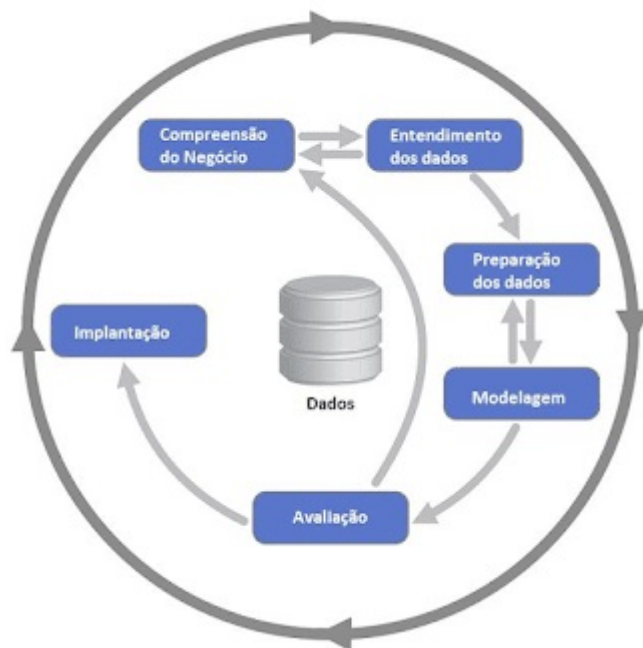
Por fim, com uma grande quantidade de dados qualificados, sistemas de *machine learning* se tornam eficazes oferecendo diversos padrões e comportamentos em que os clientes tiveram interações com os serviços da empresa. Desta maneira, é possível destacar a importância e o motivo pelo qual empresas de diferentes setores estão cada vez mais adotando soluções tecnológicas fundamentadas na IA e *machine learning* para direcionar e criar oportunidades de negócio. Abaixo, através da Figura 1 – Estratégias de dados de acordo com volume, é possível ilustrar a adoção de estratégias versus o volume de dados existentes em um determinado cenário.



**Figura 1** – Estratégias de dados de acordo com volume

Outro aspecto importante a destacar além de quantidades significativas de dados históricos é ter meios de captar novos dados para que estes, retroalimentem os sistemas de *machine learning* fazendo com que as respostas e resultados gerados estejam alinhados e atualizadas com as diretrizes de negócio e comportamento de mercado na qual a empresa está situada.

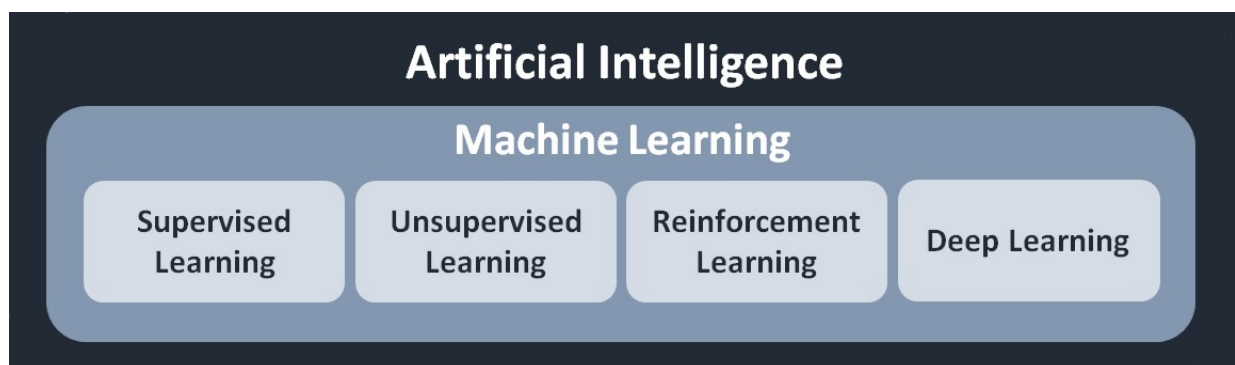
Neste sentido, esteiras de dados ou pipelines são adotados para facilitar este processo de retroalimentação. Ainda, algumas técnicas são comumente utilizadas e auxiliam na organização das etapas necessárias para construção de bons pipelines, como por exemplo, *Cross Industry Standard Process for Data Mining* (CRISP-DM), que é uma técnica/modelo de processo de mineração de dados que descreve abordagens comumente usadas por especialistas em mineração de dados para tratar e gerar soluções de dados. Logo abaixo, através da Figura 2 - Composição de Etapas Utilizadas no CRISP-DM, pode-se visualizar as etapas do fluxo de dados proposto pelo CRISP-DM.



**Figura 2** – Composição de Etapas Utilizadas no CRISP-DM

## Abordagens de *Machine Learning*

Há várias abordagens para o uso e aplicação de *machine learning* que diferem com base no tipo e volume de dados. Estas, são necessárias para melhorar a precisão dos modelos preditivos gerados. Através da Figura 3 – Categorias existentes dentro do *machine learning*, é possível identificar os 4 tipos de categorias que compõem o *machine learning*:



**Figura 3** – Categorias existentes dentro do *Machine Learning*

A fim de oferecer um entendimento sobre cada um dos 4 tipos de aprendizados de máquina, a seguir são compartilhados detalhes sobre cada um:

- **Aprendizado de máquina supervisionado (*Supervised Learning*):** O aprendizado supervisionado se inicia com um conjunto de dados e um certo entendimento de como os dados são classificados, ou seja, conhecemos a variável resposta, alvo ou dependente. Nesta categoria de aprendizado, o objetivo é encontrar padrões em dados para que possam ser aplicados em um processo analítico de tal maneira que sejam rotulados. Por exemplo, podemos utilizar o aprendizado supervisionado para classificar quanto a sua estatura entre baixa, média ou alta. Principais algoritmos: *Linear regression*; *Logistic regression*; *Decision Trees*; *Random Forest*; *Support Vector Machines (SVM)*; KNN; *Naïve-Bayes*.
- **Aprendizado de máquina não supervisionado (*Unsupervised Learning*):** O aprendizado não supervisionado é usado quando o problema lida com uma grande quantidade de dados não rotulados. Por exemplo, aplicativos de mídia social, como Twitter, Instagram e Snapchat, têm grandes quantias de dados não rotulados. Entender o significado por trás desses dados requer algoritmos que classificam os dados com base nos padrões ou *clusters* encontrados. O aprendizado não supervisionado conduz um processo iterativo, analisando dados sem intervenção humana. É usado, por exemplo, com tecnologia de detecção de *spam* por e-mail. Há muitas variáveis em e-mails legítimos e de *spam* para um analista marcar um e-mail em massa não solicitado. Em vez disso, os classificadores de *machine learning*, baseados em *cluster* e associação, são aplicados para identificar e-mails indesejados. Principais algoritmos: *Clustering & Dimensionality Reduction (SVC, PCA, K-means)*; DBSCAN; Cauterização hierárquico - acumulativo e diversivo; *Association Analysis* (Apriori; FP-Growth); *Hidden Markov Model*.
- **Aprendizado por reforço (*Reinforcement Learning*):** O aprendizado por reforço é um modelo de aprendizado comportamental. O algoritmo recebe feedback da análise de dados, orientando o usuário para o melhor resultado. O aprendizado de reforço difere de outros tipos de aprendizado, porque o sistema não é treinado com o conjunto de dados de amostra. Em vez disso, o sistema aprende por meio de tentativa e erro. Portanto, uma sequência de decisões bem-sucedidas resultará no processo sendo reforçado, porque resolve melhor o problema em questão. Principais algoritmos: *Q-Learning*; Monte Carlo *Tree Search*; *Temporal Difference*, or TD; *Asynchronous Actor-Critic Agents* (AAAC).
- **Aprendizado profundo (*Deep learning*):** *Deep learning* é um método específico de aprendizado de máquina que incorpora redes neurais em camadas sucessivas para aprender com os dados de uma maneira iterativa. *Deep learning* é especialmente útil quando você está tentando aprender padrões de dados não estruturados. Redes neurais complexas de *Deep learning* são projetadas para emular como o cérebro humano funciona, para que os computadores possam ser treinados para lidar com abstrações e problemas mal definidos. Redes neurais e *deep learning* são frequentemente usados em aplicativos de reconhecimento de imagem, fala e visão computacional. Principais algoritmos: *Convolutional Neural Networks (CNNs)*; *Long Short Term Memory Networks (LSTMs)*; *Recurrent Neural Networks (RNNs)*; *Generative Adversarial Networks (GANs)*; *Radial Basis Function Networks (RBFNs)*; *Multilayer Perceptrons (MLPs)*; *Self Organizing Maps (SOMs)*; *Deep Belief Networks (DBNs)*; *Restricted Boltzmann Machines (RBMs)*; *Autoencoders*.

## Exemplos de uso do *Machine Learning*

O uso do *machine learning* nas mais diversas aplicações só tende a crescer. Não é por capricho, mas por necessidade: muitos recursos tecnológicos que temos hoje só funcionam ou são viáveis por conta da inteligência artificial. Abaixo, compartilho alguns exemplos:

- **Combate a fraudes:** Tentativas de fraudes com cartões de crédito e outros meios de pagamentos são geradas a cada segundo no mundo todo. Felizmente, o *machine learning* tem permitido que sistemas de combate a fraudes barrem a maior parte dessas ações.
- **Recomendações de conteúdos:** Plataformas de streaming de vídeo e áudio usam o *machine learning* para analisar o histórico de conteúdo reproduzido ou rejeitado pelo usuário para fornecer a eles recomendações condizentes com os seus gostos.
- **Traduções de textos:** Traduções não podem ser feitas ao pé da letra, é necessário levar em conta contextos, expressões regionais e outros parâmetros. Graças ao *machine learning*, os tradutores automáticos estão ficando cada vez mais precisos.
- **Banco de dados autônomo:** Com auxílio do *machine learning*, bancos de dados autônomos lidam de modo automatizado com várias tarefas até então realizadas por um administrador (DBA), permitindo que o profissional cuide de outras atividades e diminuindo o risco de indisponibilidade da aplicação por falha humana.
- **Ajuste das necessidades empresariais:** O *machine learning* oferece valor potencial para as empresas que tentam alavancar grandes volumes de dados e as ajuda a entender melhor as sutis mudanças de comportamento, preferências ou satisfação do cliente. Líderes de negócios estão começando a perceber que muitas coisas que acontecem em suas empresas e indústrias não podem ser compreendidas por meio de uma consulta. Não são as perguntas que você conhece, são os padrões ocultos e anomalias enterradas nos dados que podem ajudá-lo ou prejudicá-lo.

## Materiais complementares

- Let's Code channel at Youtube: [Machine Learning além das previsões](#)
- RUSSEL, S.; NORVIG, P.; Artificial Intelligence: A Modern Approach. Prentice-Hall, Second Edition, 2003.
- FACELI, Katti; LORENA, Ana Carolina; GAMA, João; CARVALHO, André Carlos Ponce de Leon Ferreira de. Inteligência artificial: uma abordagem de aprendizado de máquina. [S.l.: s.n.], 2011.
- 5 ways AI, IoT, and security are shaping the API economy. In: MASON, Ross. 5 ways AI, IoT, and security are shaping the API economy. [S. l.]: MuleSoft, [ca. 2019]. p. 5-26.
- *Machine Learning Basics: An Illustrated Guide for Non-Technical Readers*. In: *Machine Learning Basics: An Illustrated Guide for Non-Technical Readers*. [S. l.]: Data Iku, [ca. 2019]. p. 2-27.
- *Machine Learning for Dummies*. In: HURWITZ, Judith. *Machine Learning for Dummies*. [S. l.]: John Wiley & Sons, Inc, [ca. 2018]. cap. 8-70
- BRACKETT, Michael; EARLEY, Production Susan. The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMBOK Guide). 2009.
- BAOLONG, Yang; HONG, Wu; HAODONG, Zhang. Research and application of data management based on Data Management Maturity Model (DMM) . In: Proceedings of the 2018 10th International Conference on Machine Learning and Computing. 2018. p. 157-160.
- RYU, Kyung-Seok; PARK, Joo-Seok; PARK, Jae-Hong. A data quality management maturity model. ETRI journal, v. 28, n. 2, p. 191-204, 2006

## Referências

- RUSSEL, S.; NORVIG, P.; Artificial Intelligence: A Modern Approach. Prentice-Hall, Second Edition, 2003.

- FACELI, Katti; LORENA, Ana Carolina; GAMA, João; CARVALHO, André Carlos Ponce de Leon Ferreira de. Inteligência artificial: uma abordagem de aprendizado de máquina. [S.l.: s.n.], 2011.
- 5 ways AI, IoT, and security are shaping the API economy. In: MASON, Ross. 5 ways AI, IoT, and security are shaping the API economy. [S. l.]: MuleSoft, [ca. 2019]. p. 5-26.
- *Machine Learning* Basics: An Illustrated Guide for Non-Technical Readers. In: *Machine learning* Basics: An Illustrated Guide for Non-Technical Readers. [S. l.]: Data Iku, [ca. 2019]. p. 2-27.
- *Machine Learning* for Dummies. In: HURWITZ, Judith. *Machine Learning* for Dummies. [S. l.]: John Wiley & Sons, Inc, [ca. 2018]. cap. 8-70

[Próximo Tópico >](#)