# Fundamentos Básicos do Machine Learning que você Precisa Saber

Geovana Sousa

Fonte: <https://www.learningdata.dev/post/fundamentos-basicos-do-machine-learning-que-voce-precisa-saber>

Você sabe o que é Machine Learning?

Dentre os conhecimentos principais necessários para Cientistas de Dados, temos a área de Machine Learning que é uma das bases da Inteligência Artificial aplicada à diversos problemas de negócios nas empresas.

Venha entender um pouco sobre os fundamentos básicos que você precisa entender antes de mergulhar no universo do Machine Learning no novo artigo do blog.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Ao longo dos anos a área da tecnologia passou por grandes avanços que tornaram possíveis coisas que jamais iríamos imaginar. Já pensou se você voltasse no tempo, lá para a década de 60, parasse uma pessoa qualquer na rua e dissesse que, em menos de 100 anos as pessoas carregariam consigo um aparelho que funciona como telefone, mas que o possibilita acessar todo o conhecimento do mundo com apenas alguns toques. Como você imagina que seria a sua reação?

Com certeza essa pessoa vai te achar completamente louca, assim como muitas pessoas se assustam com o fato de existir carros autônomos, sistemas de recomendações avançados, entre outras tecnologias mais avançadas. Mas agora parando para analisar, como estas tecnologias realmente funcionam? Elas são aplicações de estudos da Inteligência Artificial,como por exemplo, o Machine Learning.

Mas o que é esse tal de *Machine Learning* que tanto se ouve falar? É isso o que iremos entender neste artigo. Confira abaixo os temas que serão abordados ao decorrer do texto.

* [O que é *Machine Learning*](https://www.learningdata.dev/post/fundamentos-basicos-do-machine-learning-que-voce-precisa-saber#viewer-1eli)
* [Aprendizado Supervisionado x Não-supervisionado](https://www.learningdata.dev/post/fundamentos-basicos-do-machine-learning-que-voce-precisa-saber#viewer-b8501)
* [Aprendizado Semi-supervisionado](https://www.learningdata.dev/post/fundamentos-basicos-do-machine-learning-que-voce-precisa-saber#viewer-d9in1)
* [Aprendizado por Reforço](https://www.learningdata.dev/post/fundamentos-basicos-do-machine-learning-que-voce-precisa-saber#viewer-93k0h)
* [Conclusão](https://www.learningdata.dev/post/fundamentos-basicos-do-machine-learning-que-voce-precisa-saber#viewer-8gjl5)

## O que é Machine Learning

***Machine Learning***, traduzido para Aprendizado de Máquina, é uma área da Inteligência Artificial que tem como objetivo programar computadores para aprender a partir dos dados. Adefinição dada por Arthur Samuel faz um resumo perfeito deste conceito.

*Aprendizado de Máquina é a área de estudo que dá ao computador a habilidade de aprender sem ter sido explicitamente programado.*

*- Arthur Samuel, 1959*

Fundamentalmente, o *Machine Learning* envolve a construção de modelos matemáticos para entender um conjunto de dados. Esses modelos são conhecidos como **Algoritmos de Aprendizado de Máquina**, que são utilizados para atingir diferentes objetivos, seja classificar se um paciente tem câncer ou não, ou até mesmo, recomendar um produto de acordo com o seu padrão de visitas em um site de compras.

## Mas por que precisamos usar o Aprendizado de Máquina?

Para responder esta pergunta, vamos utilizar o exemplo do livro [***Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow***](https://www.amazon.com.br/Hands-Machine-Learning-Scikit-Learn-TensorFlow/dp/1492032646/ref=sr_1_1?dchild=1&keywords=hands+on+machine+learning&qid=1632187330&sr=8-1&ufe=app_do%3Aamzn1.fos.e05b01e0-91a7-477e-a514-15a32325a6d6), o filtro de *spam,* que tomou conta do mundo nos anos 90.

Mas primeiro, vamos relembrar como funciona este filtro. Basicamente, ele estuda os emails marcados como *spams* pelos usuários e encontra diversos padrões para futuramente, classificar novas mensagens sem o auxílio do usuário.

Agora pense comigo como seria construído este filtro sem a utilização do *Machine Learning.* A princípio, o desenvolvedor conseguiria encontrar algumas palavras-chaves que estão presentes na grande maioria desses emails, e até alguns outros padrões de mensagem. Em seguida, ele construiria um algoritmo para detectar esses padrões.

Você consegue enxergar qual o problema desta solução?

Ao fazer isto, você acaba criando um loop eterno de programação. Todas as vezes que criarem diferentes tipos de *spams,* o desenvolvedor precisará refazer todo o processo à mão,o que gera muito trabalho. E por isso, precisamos que o computador faça esta tarefa.

## Aprendizado Supervisionado x Não-supervisionado

Dentre os diferentes tipos de classificações dos algoritmos de *Machine Learning*, temos aqueles que são categorizados a partir da quantidade e do tipo de supervisão que recebem durante o período de treino. Entre estes, temos quatro categorias de aprendizado: supervisionado, não-supervisionado, semi-supervisionado e o por reforço.

### Aprendizado Supervisionado

Neste tipo de aprendizado, o dataset de treinamento que o algoritmo recebe já vem com a sua classificação, sua *label.* Os algoritmos desta categoria são treinados para classificar dados ou prever valores númericos, de tal forma que o modelo se ajusta aos *inputs* por meio de *cross validation,* e desta forma atingem os melhores resultados.

Basicamente, o aprendizado supervisionado usa um dataset de treinamento que inclui os *inputs* e *outputs* corretos para ensinar o modelo com o tempo. Em seguida, o algoritmo mede sua acurácia por meio da função de perda (*loss function*), e ajustando até que o erro seja minimizado.

Diagrama

Descrição gerada automaticamente

Esta categoria é separada entre dois tipos de problemas, classificação e regressão.

* **Classificação:** utiliza algoritmos para distribuir dados em categorias específicas, observando padrões de acordo com os *labels* já existentes. Os principais algoritmos de classificação são: classificadores lineares, *Support Vector Machines (SVM)*, árvores de decisão, *k-nearest neighbor* e *random forest.*
* **Regressão:** utilizado para entender a relação entre variáveis dependentes e independentes, normalmente tem o objetivo de prever resultados com base nos dadosde teste. Os principais algoritmos são: regressão linear e regressão polinomial.

Algoritmos importantes do Aprendizado Supervisionado

* *k-Nearest Neighbors*
* Regressão Linear
* Regressão Logística
* *Support Vector Machines (SVM)*
* Árvores de Decisão e *Random Forests*
* *Neural Networks*

## Aprendizado Não-supervisionado

Pela definição do aprendizado anterior, você pode supor que o não-supervisionado seria com dados não categorizados. Se você fez esta suposição, está totalmente correto!

No aprendizado não-supervisionado, o algoritmo tenta identificar padrões sem resultados pré-definidos. Mas como isso funciona? Bem, vamos supor que você forneça ao algoritmo várias fotos de diferentes frutas. A partir destas informações, o algoritmo tentará encontrar padrões nas imagens e criará algumas divisões de categorias de acordo com a divisão dos padrões encontrados.

Diagrama, Esquemático

Descrição gerada automaticamente

Algorimos importantes do Aprendizado Não-supervisionado

Clustering:

* K-médias
* DBSCAN
* Hierarchical Cluster Analysis (HCA)

Detecção de Anomalia e Detecção de Novidades

* One-class SVM
* Isolation Forest

Visualização e Redução de Dimensionalidade

* Análise de Componentes Principais (PCA)
* Kernel PCA
* Locally Linear Embedding (LLE)
* t-Distributed Stochastic Neighbor Embedding (t-SNE)

Regras de Associação

* Apriori
* Eclat

## Aprendizado Semi-Supervisionado

Alguns algoritmos conseguem lidar com dados classificados e não classificados juntos. Isso porque, rotular todos os dados gasta muito tempo e acaba tendo um custo muito alto, e para que não se tornasse um problema, foram criados os algoritmos de aprendizado semi-supervisionado.

A maioria destes algoritmos são uma mistura de supervisionados e não-supervisionados, e são muito utilizados em modelos de reconhecimento facial, por exemplo, o Google Photos que consegue reconhecer pessoas nas suas fotos depois que você informa algumas vezes quem elas são.

## Aprendizado por Reforço

Diferentemente dos aprendizados que vimos até agora, o aprendizado por reforço é um sistema onde o agente inteligente aprende a atingir objetivos por meio de recompesas estabelecidas para o modelo.

Basicamente, o agente é inserido em um ambiente, e ele deve aprender por tentativas de ações que serão recompensadas ou penalizadas, tendo como objetivo adquirir a maior pontuação possível. Para ficar mais fácil, podemos pensar em como animais são adestrados. Primeiro o treinador vai acostumar o animal a obedecer suas ordens, definindo quais são as regras que ele deve seguir. Em seguida, o animal começará a praticar com o objetivo de receber algum tipo de agrado, um petisco talvez, e com o tempo ele entenderá que todas as vezes em que fizer aquela ação receberá uma recompensa, e ao contrário será penalizado.

Assista o [vídeo abaixo](https://youtu.be/1aEXbyHB3_k) para ver um exemplo da aplicação deste conceito com uma galinha.

Vídeo: <https://youtu.be/1aEXbyHB3_k>

## Conclusão

Chegamos ao final de mais um artigo no blog, onde aprendemos sobre uma das ferramentas principais no universo da ciência de dados, o ***Machine Learning***. Assim como as diferenças entre os principais modelos e algoritmos que formam o portfólio do aprendizado de máquina.

----------------------------------------------------------------------------------------------------------

**Espero que tenham gostado do conteúdo. Se desejam aprender mais sobre *Data Science*, *Machine Learning*, Inteligência Artificial e outros assuntos relacionados, me sigam no LinkedIn e no Instagram para não perderem as próximas atualizações.**😉