

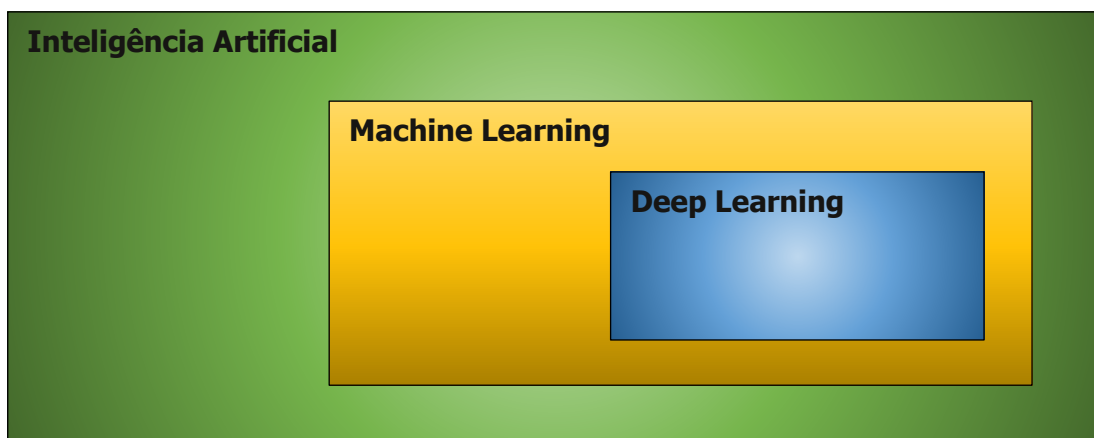
## Introdução a Ciência de Dados

A ciência de dados é a área que mais cresce atualmente porque geramos dados a todo segundo e por isso as empresas antenadas passaram a valorizar os dados gerando valor para si. O uso dos dados pelas empresas também é conhecido como Cultura *Data Driven*, ou seja, é quando a empresa usa os dados como ferramenta para tomada de decisões mais assertivas. Então, com a necessidade de extrair as informações essenciais dos dados emergiu a profissão de cientista de dados, onde esse profissional aplica técnicas para produzir informações para melhores tomadas de decisões e geração de produtos e serviços inovadores.

A ciência de dados é uma área multidisciplinar abrangendo diferentes campos de conhecimento: estatística, matemática, métodos científicos, inteligência artificial (IA), análise de dados, etc. Então todo e qualquer conhecimento que você já tenha, seja advindo das suas experiências escolares, universitárias, profissionais e/ou simplesmente sua experiência de vida, cabem nessa nova profissão!

## Ciência de Dados x Machine Learning

Como vimos acima a ciência de dados é um campo de estudo com outras disciplinas. Já o famoso *Machine Learning* (ML) é uma subárea da Inteligência Artificial (IA) que usa algoritmos para criação de modelos de aprendizagem, ou seja, modelos que aprendem a gerar novos resultados baseados no que eles foram treinados para realizar. Além do ML, também temos o *Deep Learning* (DL) que por sua vez é uma subárea do ML e se caracteriza pelo uso de redes neurais para resolver problemas mais complexos. Lembrando que a IA é aquela área em que máquinas e soluções inteligentes são criadas bem como: robôs, carros e casas inteligentes, etc.



## Exemplos de aplicações reais de ML/DL

Identificação de spam e malwares, detecção de fraudes online, reconhecimento de imagem, diagnósticos médicos, assistentes virtuais, carros inteligentes, sistemas de recomendações, etc. Grandes empresas como Google, Facebook, Amazon, Netflix são ótimos exemplos de desenvolvedores de sistemas inteligentes empregando ML/DL.

Como podemos perceber estamos em constante contato com aplicações reais de modelos de ML/DL no nosso dia a dia!

## O trabalho do Cientista de Dados

O trabalho do cientista de dados contempla algumas etapas essenciais seja para desenvolvimento de modelos de ML ou DL, projetos de pesquisa ou trabalhos requeridos por empresas.

Podemos dividir o workflow nas seguintes etapas:

1. *Definição do problema:* entendimento do trabalho a ser realizado.
2. *Aquisição de dados:* obtenção dos dados essenciais para o trabalho.
3. *Preparação dos dados:* tratamento dos dados para posterior análises.
4. *Análise exploratória dos dados:* entendimento da relação entre as variáveis.
5. *Criação de modelos de ML/DL:* implementação e teste de diferentes modelos.
6. *Apresentação dos resultados:* comunicação dos principais resultados e achados.
7. *Deploy do modelo:* produção e monitoramento do modelo.

Todas essas etapas se comunicam e criam um trabalho cíclico em que podemos ir e voltar de acordo com a necessidade do trabalho caracterizando um trabalho de intensa pesquisa e experimentação. Vale destacar também que as etapas 3 e 4 representam a maior demanda de esforço e trabalho, pois a preparação e análise dos dados já fornecem por si só um primeiro resultado. Além disso, é a partir de tais etapas que entendemos os dados, trabalhamos as variáveis, testamos hipóteses e definimos os alvos para criação subsequentes dos modelos.

## **Objetos de entrega de um Cientista de Dados**

É muito importante sempre ter em mente que o Cientista de Dados não é nenhum mágico ou carta coringa que surgirá sempre com a solução para todo e qualquer problema! Por tanto, é essencial conhecermos bem quais tipos de entregas somos capazes de oferecer para solucionar determinado tipo de problema.

Basicamente, são três os objetos de entrega do cientista de dados de acordo com a proposta do trabalho, ou seja, problema a ser resolvido:

- 1) *Problema de análises*: são aqueles em que o trabalho consiste somente na análise exploratória dos dados e entrega desses resultados.
- 2) *Problema de regressão*: são aqueles em que é necessário criar um modelo capaz de prever uma variável contínua quantitativa, ou seja, um valor mensurável.
- 3) *Problema de classificação*: são aqueles em que um modelo capaz de classificar, ou seja, atribuir um dado a uma determinada categoria é criado.

Considere os seguintes exemplos:

- a) Previsão do comportamento do cliente, ele continuará cliente ou não?
- b) Precificação de produtos de acordo com várias características: custo de produção, estoque, saída, avaliação do cliente, etc.
- c) Personalização de ofertas, quais produtos estão encalhados? Quais tiveram maior saída? Quais dias teve maior fluxo de clientes? Etc.

Agora atribua a cada um deles um dos 3 tipos de problemas estudados e entregues por um cientista de dados!

## **Boas práticas de trabalho**

Como práticas essenciais para um bom desempenho de trabalho devemos sempre ter e manter o foco do trabalho, além disso buscar ferramentas adequadas para a execução do mesmo. Também, a comunicação e documentação contínua do trabalho traz ganho tanto para o cientista quanto para a equipe, afinal a troca de conhecimento agrega valor a qualquer trabalho. Quanto aos algoritmos implementados é de suma importância que seja usado algum

método de versionamento para que as mudanças sejam sempre computadas e nunca perdidas num todo. Um ambiente organizado e bem estruturado garantirá acima de tudo a realização de um trabalho limpo e pronto para ser compartilhado.

### **DICAS IMPORTANTES**

Além de conhecimentos básicos de programação (principalmente em python), estatística e matemática, a Ciência de Dados requer toda sua sagacidade e criatividade para resolver problemas. Afinal, como cientistas devemos sempre ser capazes de propor soluções simples para problemas de diferentes níveis de complexidade.

Lembre-se que nem tudo na Ciência de Dados se resolve com Machine Learning/Deep Learning, pois as vezes o trabalho requer apenas uma análise profunda dos dados. Como isso percebemos que a Ciência de Dados trata da solução de problemas enquanto o ML/DL refere-se a algoritmos (técnicas) aplicados a solução desses problemas.

O ciclo de vida dos projetos/trabalhos do cientista de dados não é linear, portanto, as etapas de comunicam e requerem intensa pesquisa e experimentação.

Por fim, lembre-se sempre de que a sua atitude é que governa até onde você é capaz de chegar! E como fonte de inspiração temos a interpretação do conceito de inteligência feito por grandes personalidades:

**Albert Einstein:** "The true sign of intelligence is not knowledge but imagination."

**Pablo Picasso:** "I am always doing that which I cannot do, in order that I may learn how to do it."

**Daniel Kahneman:** "Intelligence is not only the ability to reason; it is also the ability to find relevant material in memory and to deploy attention when needed."

Você pode ler um pouco mais sobre esses e outras interpretações no seguinte artigo: [Intelligence Explained By Most Intelligent People | by Diego Salinas | Cognitive AI | Medium](#)