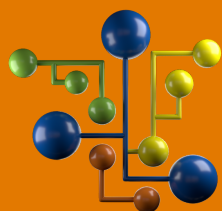




# Guia de Carreira Engenheiro de Machine Learning



**Data Science  
Academy**

Equipe Data Science Academy  
[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)



# Conteúdo do Guia

► Acompanhe o descritivo dos conteúdos deste Guia:

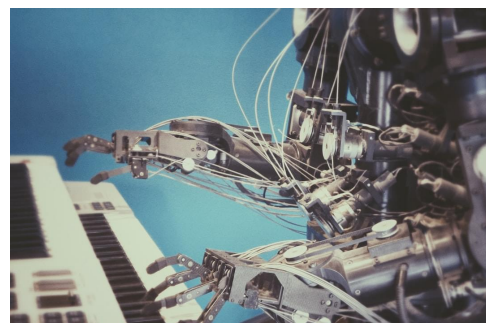
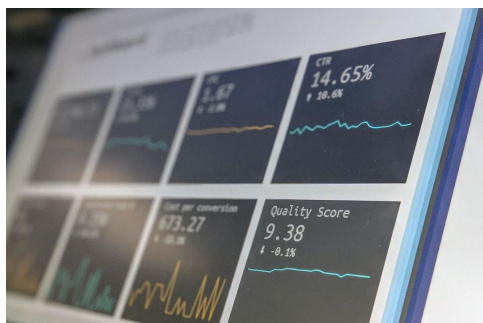
Página	Descritivo
1	Capa
2	Conteúdo do Guia
3	Bem-vindo / Por Que Precisamos de Machine Learning?
4	O Que Exatamente Significa Treinar um Modelo de Machine Learning
5	Quem é o Engenheiro de Machine Learning?
6	Por Onde Começar em 6 Passos
7,8 e 9	Aprenda a Programar em Python e/ou C ++
10	Aprenda a Construir, Treinar e Avaliar Modelos de Machine Learning
11	Compreenda os Fundamentos de Engenharia de Software
12	Desenvolva Suas Habilidades em Cloud Computing
13	Seja Proficiente em AutoML
14	Aprenda a Fazer o Deploy de Modelos de Machine Learning
15	Engenheiros de Machine Learning Estão em Alta Demanda
16	Como se Preparar?
17	Créditos



# Bem-vindo(a)!

Ansioso por desbravar o universo da Ciência de Dados e não sabe por onde começar? Nós ajudaremos você. Preparamos um guia que vai ajudar você a compreender o que faz um **Engenheiro e Machine Learning** e como iniciar sua preparação! Confira.

## Por Que Precisamos de Machine Learning?



Precisamos de aprendizado de máquina porque queremos automatizar certos processos e resolver problemas de negócio. O aprendizado de máquina nasceu do reconhecimento de padrões e da ideia de que os computadores podem aprender sem serem programados explicitamente para realizar tarefas específicas. Pesquisadores interessados em Inteligência Artificial ansiavam por ver se os computadores poderiam aprender com os dados.

O aspecto iterativo do aprendizado de máquina é importante porque, à medida que os modelos são expostos a novos dados, eles podem se adaptar de forma independente. Eles aprendem com cálculos anteriores para produzir decisões e resultados confiáveis. É uma ciência que não é nova, mas que ganhou um novo impulso graças ao Big Data e ao aumento da capacidade computacional.

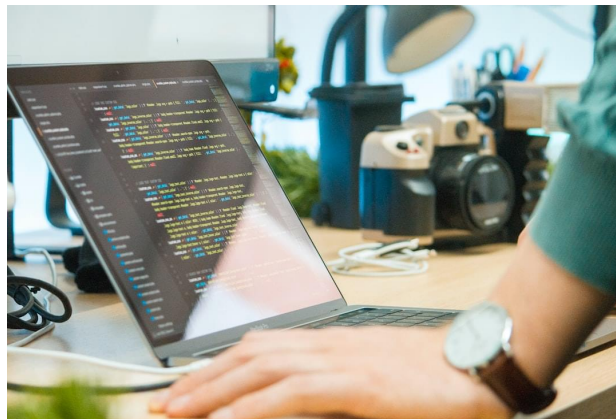
Embora muitos algoritmos de aprendizado de máquina existam há muito tempo, a capacidade de aplicar automaticamente cálculos matemáticos complexos a Big Data – repetidamente, cada vez mais rápido – é um desenvolvimento recente, com pouco mais de uma década. Logo, há muito espaço para crescimento ao longo dos próximos anos.



# O Que Exatamente Significa Treinar um Modelo de Machine Learning?

Machine Learning (ML) é uma forma de Inteligência Artificial (IA) que permite que um sistema aprenda com dados ao invés de programação explícita.

Depois que um algoritmo de ML é escrito, ele deve ser “treinado” antes de ser implantado em seu uso pretendido. O treinamento é o processo pelo qual a máquina aprende. Algoritmos são alimentados com dados de treinamento, possibilitando a produção de modelos. O modelo aprende o padrão nos dados (se o padrão existir) e ao receber novos dados, é capaz de detectar o padrão e fazer previsões.



Um modelo de aprendizado de máquina é a saída gerada depois que um algoritmo de aprendizado de máquina é treinado com ingestão de dados. Uma vez treinado, quando um modelo de aprendizado de máquina é alimentado com novos dados, ele produz uma saída. Quando o modelo preditivo é alimentado com novos dados, ele faz uma previsão com base no padrão nos dados que treinaram o modelo.

Por meio de seu treinamento e aprendizado online iterativo, um modelo de aprendizado de máquina pode melhorar muito sua compreensão dos tipos de associações que existem entre os elementos de dados. Devido à sua complexidade e tamanho, esses padrões e associações seriam facilmente ignorados pela observação humana. Técnicas de aprendizado de máquina são necessárias para melhorar a precisão das análises. Dependendo da natureza do problema de negócio, existem diferentes abordagens com base no tipo e no volume dos dados.



## Quem é o Engenheiro de Machine Learning?

Um Engenheiro de Machine Learning deve entender os elementos listados na seção anterior, bem como em quais situações aplicá-los. As quatro abordagens básicas utilizadas são aprendizado supervisionado, aprendizado não supervisionado, aprendizado por reforço e aprendizado profundo.



As diferenças entre essas abordagens estão nos dados a serem usados para criar um modelo de Machine Learning.

O aprendizado supervisionado é usado com dados já fornecidos com rótulos, por exemplo, identificação de animais a partir de novas imagens.

O aprendizado não supervisionado é usado com dados não rotulados e geralmente é usado para identificar e-mails de spam ou segmentar clientes por similaridade.

O aprendizado por reforço é semelhante ao aprendizado supervisionado, pois usa dados rotulados. No entanto, o aprendizado por reforço é feito sem o benefício dos dados de treinamento, melhorando sua modelagem por tentativa e erro de dados do mundo real.

O aprendizado profundo (**Deep Learning**) incorpora redes neurais para aprender com os dados de maneira iterativa. É especialmente útil no aprendizado de padrões de dados não estruturados em aplicativos de Processamento de Linguagem Natural e Visão Computacional. Há modelos de Deep Learning para aprendizado supervisionado, não supervisionado e por reforço.

Um Engenheiro de Machine Learning deve utilizar técnicas e algoritmos avançados de programação para criar um sistema capaz de ingerir um tipo específico de dados e transformá-lo na saída da modelagem desejada.



# Por Onde Começar em 6 Passos

Agora que você tem uma boa visão sobre o que é e o que faz um Engenheiro de Machine Learning e decidiu que deseja trabalhar nessa função, aqui estão os 6 passos por onde começar.



**Passo 1 – Aprenda a Programar em Python e/ou C++**



**Passo 2 -Aprenda a Construir, Treinar e Avaliar Modelos de Machine Learning**



**Passo 3 – Compreenda os Fundamentos de Engenharia de Software**



**Passo 4 – Desenvolva Suas Habilidades em Cloud Computing**



**Passo 5 – Seja Proficiente em AutoML**

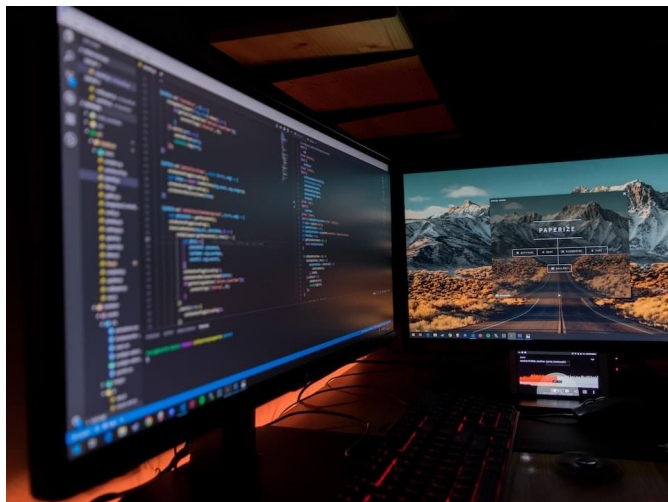


**Passo 6 – Aprenda a Fazer o Deploy de Modelos de Machine Learning**





# 1- Aprenda a Programar em Python e/ou C++



O trabalho de um Engenheiro de Machine Learning envolve programação e você precisará demonstrar proficiência em Python e/ou C++ e suas bibliotecas associadas. Python é amplamente usada no mercado, sendo mandatório dominar a linguagem e como usá-la para criar modelos de Machine Learning.

Mas a Linguagem Python pode apresentar problemas de performance e por isso dominar uma linguagem como C++ pode ser decisivo para construir modelos de alta performance e integrá-los em diferentes tipos de aplicações. Se está começando, comece com a Linguagem Python e mais tarde inclua C++ na sua caixa de ferramentas. O curso gratuito **Python Fundamentos Para Análise de Dados** pode ser um bom ponto de partida.

Também é uma boa ideia se familiarizar com as bibliotecas TensorFlow e PyTorch, as duas mais amplamente usadas no mercado atualmente. Ambas são executadas em CPUs, GPUs e outros tipos de processadores. Para modelos mais simples, Scikit-Learn também pode ser uma boa opção. Há ainda bibliotecas para linguagens C++, Scala, Java e R, dependendo de qual linguagem será usada na construção dos modelos.

Para iniciantes, recomendamos o PyTorch, um framework que pode ser usado com o modelo de programação imperativo familiar aos desenvolvedores. Ele permite que Engenheiros de Machine Learning usem instruções padrão da Linguagem Python que podem ser usadas para implementar redes neurais profundas.



Aqui estão algumas outras linguagens de programação que você pode considerar aprender para uma carreira como Engenheiro de Machine Learning:

## Linguagem

Uma linguagem estatística gratuita e de código aberto que foi lançada em 1995 como descendente da linguagem de programação S. A linguagem R oferece uma gama de pacotes de alta qualidade, específicos de domínio para atender a quase todos os aplicativos estatísticos e de visualização de dados incluindo redes neurais, regressão não linear, plotagem avançada e muito mais. Embora seja mais trabalhoso fazer o deploy de modelos criados em R, isso pode ser necessário dependendo do tipo de projeto.

## Linguagem

Significando “Structured Query Language”, SQL está no centro do armazenamento e recuperação de dados há décadas. SQL é uma linguagem específica de domínio para gerenciar dados em bancos de dados relacionais e é uma habilidade obrigatória para quem deseja trabalhar com dados e que confiam em SQL para atualizar, consultar, editar e manipular bancos de dados e extrair dados.

## Linguagem

Uma das linguagens de uso geral mais antigas usadas no mercado, a força do Java está em parte em sua popularidade e onipresença: muitas empresas, especialmente grandes empresas internacionais, usam Java para criar sistemas e aplicativos de back-end para desktop, dispositivos móveis ou web. Muitas aplicações de missão crítica são desenvolvidas em Java ou C++.





## Linguagem Scala

Fácil de usar e flexível, Scala é a linguagem de programação ideal para lidar com grandes volumes de dados. Combinando programação orientada a objetos e funcional, Scala evita bugs em aplicativos complexos com seus tipos estáticos, facilita o processamento paralelo em larga escala e, quando usada com o Apache Spark, fornece computação em cluster de alto desempenho.

## Linguagem julia

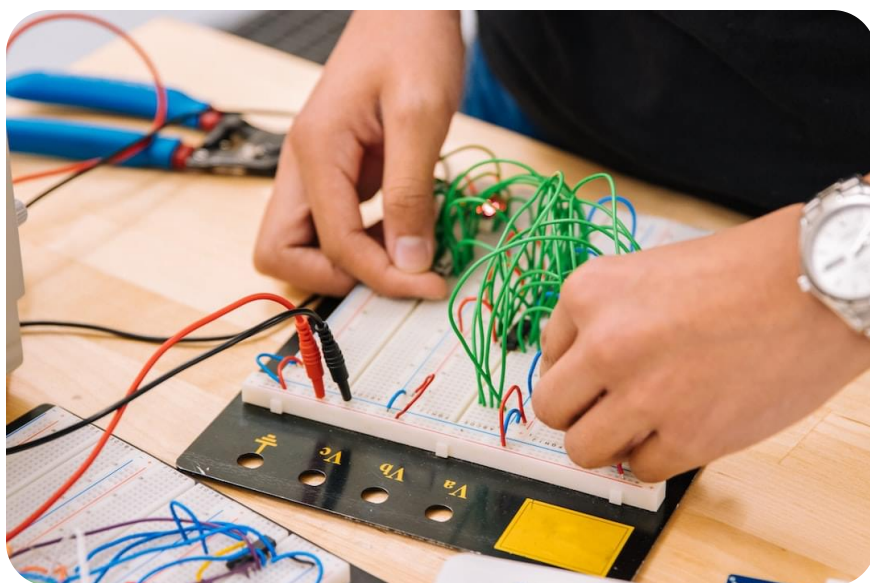
Uma linguagem de programação muito mais nova do que outras nesta lista, Julia, no entanto, causou uma boa impressão rapidamente graças ao seu desempenho, simplicidade e legibilidade extremamente rápidos. Projetado para análise numérica e ciência computacional, Julia é especialmente útil para resolver operações matemáticas complexas, o que explica por que está se tornando comum no setor financeiro. Também está se tornando amplamente conhecida como uma linguagem para Inteligência Artificial, e muitos grandes bancos agora estão usando Julia para análise de risco.

**Todas as linguagens acima são estudadas em diversos cursos na DSA. Confira clicando [aqui](#).**



## 2- Aprenda a Construir, Treinar e Avaliar Modelos de Machine Learning

Normalmente é o Cientista de Dados o profissional responsável por construir, treinar e avaliar modelos de Machine Learning. Mas Cientistas de Dados fazem isso sob a ótica da ciência, como um experimento visando buscar o melhor modelo possível para resolver determinado problema de negócio.



O Engenheiro de Machine Learning também deve saber construir, treinar e avaliar modelos de Machine Learning, mas não com o objetivo de desenvolver um experimento e sim de construir um software capaz de fazer previsões. Se o Cientista de Dados tem o foco em ciência, o Engenheiro de Machine Learning deve ter o foco mais comercial, sempre buscando a melhor forma de otimizar a performance do modelo para melhorar a experiência do usuário final que trabalhará com a aplicação baseada em Machine Learning.

Logo, dominar o processo de construção de modelos de Machine Learning é fundamental.



## 3- Compreenda os Fundamentos de Engenharia de Software



O Engenheiro de Machine Learning tem como função principal construir soluções de software baseadas em Machine Learning. Mas construir um software não é algo trivial e requer conhecimentos que vão desde arquitetura até padrões de desenvolvimento. Esse conhecimento será importante para saber a melhor forma de acoplar um modelo de Machine Learning em uma aplicação que realizará diversas outras tarefas como controle de sessão, autenticação, consulta a banco de dados, envio de e-mails, etc... Machine Learning é apenas um “pedaço” de um software maior e compreender como incluir esse “pedaço” em um software requer conhecimentos em engenharia de software.

Aliás, essa é uma habilidade crítica nos dias de hoje, uma vez que o mundo já é governado por software.



## 4- Desenvolva Suas Habilidades em Cloud Computing

Todos estão indo para a nuvem.

E você?



O crescimento de Cloud Computing nos últimos anos mostra a consolidação dessa nova filosofia de trabalho com tecnologia. Para o Engenheiro de Machine Learning há diversas ferramentas em nuvem que facilitam ou agilizam o trabalho, como AWS SageMaker ou Microsoft Azure Data Factory.

Recomendamos aqui dominar primeiro um provedor em nuvem e a AWS é uma boa opção para os iniciantes. Por ser a líder do mercado, é a que oferece o maior número de serviços em nuvem. Aqui você encontra algumas das ofertas da AWS para MLOps (Operação de Machine Learning): **AWS MLOps**.

Outra vantagem em dominar um provedor de Cloud Computing é a chance de obter as certificações que muitos deles oferecem, demonstrando para o mercado o seu conhecimento e se diferenciando da maioria.





## 5- Seja Proficiente em AutoML

AutoML (ou Automated Machine Learning) é o uso de ferramentas para automatizar a construção, treinamento e avaliação de modelos de Machine Learning. É possível testar diversas combinações de diferentes algoritmos e hiperparâmetros, a fim de encontrar o melhor modelo possível.

O conceito não é novo e AutoML pode ser feito 100% via programação. Mas o conceito ganhou popularidade com frameworks que automatizam as tarefas e facilitam o trabalho.

É sempre bom lembrar que seu trabalho não é construir modelos. Seu trabalho é resolver problemas de negócio. A construção de modelos é parte da solução. Se podemos automatizar essa parte e dedicar mais tempo em outras tarefas, devemos sempre considerar o AutoML em nosso dia a dia.

Mas AutoML não é a solução para todos os problemas. Além de diversas limitações com relação aos algoritmos que são testados, o resultado pode dificultar a interpretabilidade do modelo, o que pode ser crucial para determinados projetos. E por isso compreender bem o processo de construção de modelos é fundamental.

**Dominar AutoML é um requisito básico para Engenheiros de Machine Learning e recomendamos começar com um dos melhores frameworks para esse fim, o H2O.**



## 6 - Aprenda a Fazer o Deploy de Modelos de Machine Learning



Deploy de modelos de Machine Learning é o processo de publicar o modelo após o treinamento e permití-lo fazer as previsões, objetivo para o qual o modelo foi treinado.

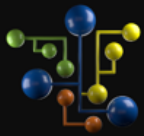
Há inúmeras formas e ferramentas para fazer o deploy de um modelo. Mas não é só isso. Após o deploy, é preciso cuidar do versionamento, já que modelos de Machine Learning serão frequentemente re-treinados e atualizados.

É preciso ainda cuidar da forma como o modelo vai receber os dados. O treinamento de um modelo ocorre em um ambiente controlado, onde os dados são pré-processados antes do treinamento. O mesmo pré-processamento aplicado aos dados de treino deve ser aplicado aos novos dados.

Esse fluxo deve ser construído e mantido. Aprender esse processo é a chave do trabalho de um Engenheiro de Machine Learning que deve ser capaz de construir pipelines de dados e interagir com Engenheiros de Dados para obter os dados da forma esperada por cada modelo em produção.



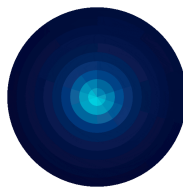
## *Formação Engenheiro de Machine Learning*



Data Science  
Academy

**Engenheiros de Machine Learning estão em alta demanda e este é o momento ideal para entrar nessa carreira.**

E a DSA, no seu compromisso de oferecer somente cursos de alto nível, vai ajudar a você a se tornar um Engenheiro de Machine Learning. Clique no botão abaixo e conheça o programa completo de capacitação, 100% online e 100% em português, que prepara Engenheiros de Machine Learning em nível profissional.

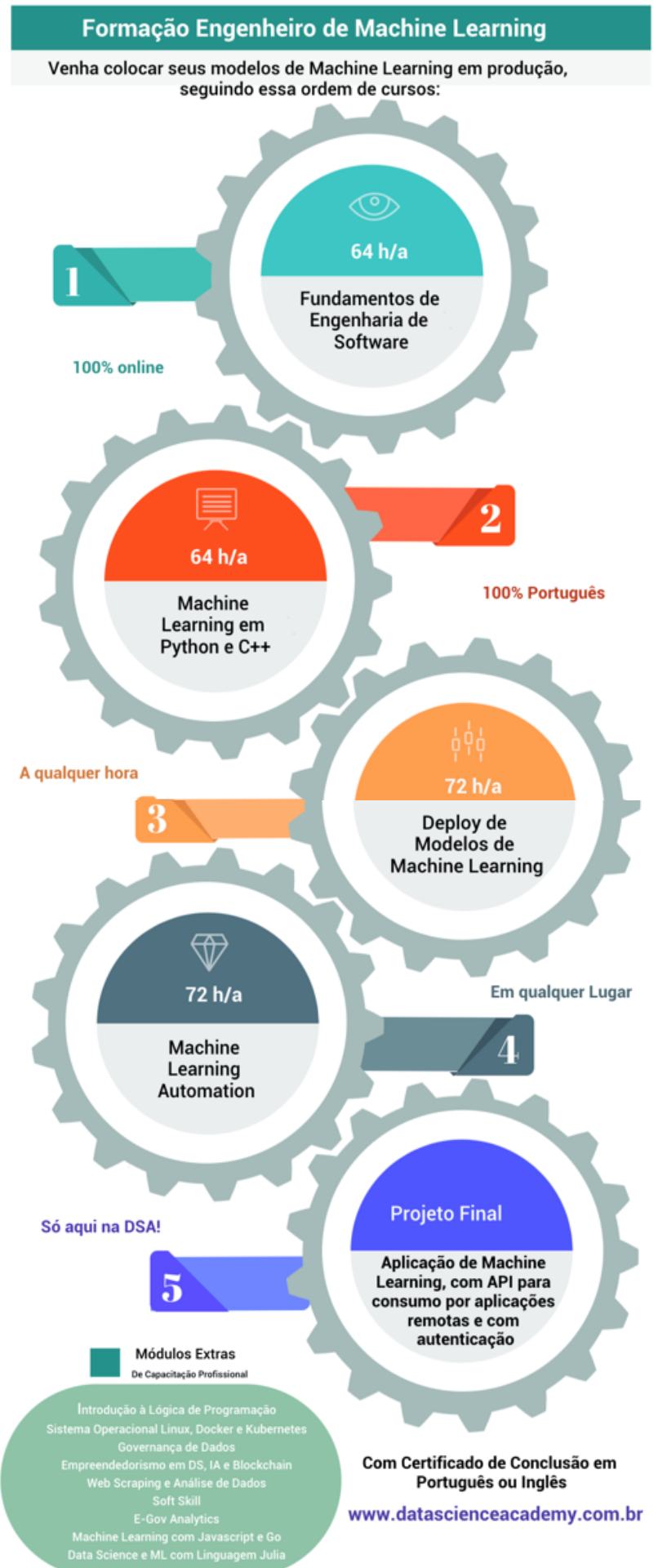


**Formação Engenheiro de Machine Learning**





## Como se Preparar?





# Créditos

## Equipe Data Science Academy

- Equipe DSA, 2022, Guia de Carreira em Engenharia de Machine Learning
- Versão 1.0
- Site [www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)