**IMERSÃO DADOS**

Aula 5: Machine Learning, Sci-kit learning e desafios envolvidos.

## **PARTICIPE E VÁ MAIS FUNDO**

Seja bem-vindo e bem-vinda à última aula da Imersão Dados 3ª Edição!

Nesta aula, finalmente vamos criar o nosso **primeiro** modelo de **Machine Learning** para classificar um novo composto com o intuito de prever se o MoA foi ativado ou não. Para isso, vamos usar a famosa biblioteca [scikit-learn](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html" \t "_blank) para testar a **regressão logística** e entender o funcionamento de uma **árvore de decisão** (AD). Além disso, discutiremos problemas **importantes** como o **overfit**.

Para ver os desafios e todo o conteúdo que foi dado em aula siga os seguintes passos:

* Acesse o notebook da [Aula 05](https://drive.google.com/file/d/1VhVP7Imj2fXDvvj4JDKYXUk9gCIV9RrQ/view?usp=sharing).
* Na parte superior esquerda, clique em >File, logo depois em >Save a copy in Drive.
* Se você não estiver logado em uma conta **gmail**, um pop-up solicitará que você crie ou faça login em uma conta google.
* Feito o login, uma **cópia** da aula é criada em seu drive (pasta Colab Notebook, criada automaticamente).
* Abra o notebook e boa diversão.

Neste link, temos o Github do projeto onde está localizado a [base de dados e os notebooks das aulas](https://github.com/alura-cursos/imersaodados3).

As **respostas para os desafios da aula 04** você encontra neste [notebook](https://drive.google.com/file/d/1OBMzDtg3c38TNNxOgTXOhe-qTXFn_A3e/view?usp=sharing) e da aula 05 neste [notebook](https://drive.google.com/file/d/1Uewkts4jc4wp7VBczPzlNIz1UGRiv5Hk/view?usp=sharing), a maioria das soluções utilizaram conceitos ensinados em aula, entretanto é importante salientar que existem diversas maneiras de resolver os exercícios.

Se tiver dificuldades ao utilizar o Colab, acesse este [artigo da Alura](https://www.alura.com.br/artigos/google-colab-o-que-e-e-como-usar) para entendê-lo.

Faça os **desafios deixados no final de cada aula** e não deixe de compartilhar seus resultados nas redes sociais marcando os instrutores e adicionando a **#ImersaoDados**

E não esqueça, agora no repositório do nosso [projeto](https://github.com/alura-cursos/imersaodados3) criamos a vitrine vai lá conferir os projetos incríveis que vocês estão desenvolvendo. E para ter seu projeto nesta vitrine, a única coisa que precisa é no repositório do seu projeto adicionar os tópicos alura e imersão dados. Para fazer isso é só clicar na engrenagem de configurações na parte superior direita do seu repositório ao lado de About.

Eu e toda equipe estamos presentes no nosso espaço de discussão dentro do Discord, não deixe de tirar suas dúvidas e participar da comunidade.

Esse projeto foi inspirado em um desafio do [Laboratory innovation science at Harvard](https://lish.harvard.edu/" \t "_blank) disponibilizando os dados em uma competição no [kaggle](https://www.kaggle.com/c/lish-moa" \t "_blank).

## Projeto final

Agora sim chegou o momento de se dedicar ao projeto final são **10 bolsas de estudo de 100% do Bootcamp Data Science Aplicada**, todas as informações sobre o projeto e a premiação você vai encontrar [neste repositório](https://github.com/alura-cursos/imersao-dados-desafio-final).

E se tiver qualquer dúvida na hora de criar o seu projeto no github, você pode assistir este [vídeo](https://www.youtube.com/watch?v=pn9hLVuIuPI).

## Links extras bacanas

Para dar um mergulho ainda mais profundo, dê uma olhada nos links abaixo:

* Assista esse [vídeo](https://www.youtube.com/watch?v=K6HmoMnqeI8) para entender mais sobre o Overfit
* Quer entender como funciona o [Machine Learning](https://www.youtube.com/watch?v=CAu97npy7zQ" \t "_blank)?
* Que tal conhecer como é trabalhar com [Machine Learning na Europa](https://www.youtube.com/watch?v=dFLrwi4Oz_o&t=1s" \t "_blank)?
* [Otimização de hiperparâmetros](https://www.alura.com.br/artigos/otimizacao-de-parametros?utm_source=gnarus&utm_medium=timeline)

## Compartilhe seus resultados!

É muito legal ver seus resultados e isso também é muito importante para sua carreira! Para isso, você pode escrever artigos no Linkedin e no Medium com link para seu Colab, texto comentando o que fez de diferente e marcando nós, instrutores e também a Alura (@aluraonline). Dá até para fazer um vídeo e encher a gente de orgulho! :)