

# Projeto de Data Science

## Data Science

---

**NOME:** Flávia Roseane Alves de Souza **CÓDIGO:** 198726  
**NOME:** Vinícius do Carmo Loeblein **CÓDIGO:** 198779

## Tema do Projeto

Este projeto tem como objetivo investigar a relação entre o uso de redes sociais e a produtividade dos indivíduos. O tema é relevante no contexto atual, onde o uso de plataformas digitais está cada vez mais presente no cotidiano das pessoas, potencialmente afetando seu desempenho profissional e acadêmico.

## URL do GitHub

O projeto está hospedado no GitHub no seguinte repositório:  
<https://github.com/Flaviasoz/social-media-productivity> (acesso concedido ao usuário @diegopat)

## Dataset Utilizado

### Origem

O dataset foi obtido do Kaggle: **Social Media vs Productivity**:  
<https://www.kaggle.com/datasets/mahdimashayekhi/social-media-vs-productivity>

### Variáveis Principais

1. **Idade:** Faixa etária dos respondentes.
2. **Tempo em redes sociais:** Horas diárias gastas em plataformas sociais.
3. **Produtividade autoavaliada:** Escala de 1 a 100.
4. **Horas de sono:** Duração média do sono por noite.
5. **Nível de estresse:** Auto Avaliação em escala numérica.

## Transformações Realizadas

1. Tratamento de valores faltantes.

## Bibliotecas Utilizadas

1. Pandas (manipulação de dados)
2. NumPy (operações numéricas)
3. Matplotlib/Seaborn (visualização)
4. Scikit-learn (modelos de machine learning)
5. Jupyter (notebooks interativos)

## Modelos Utilizados

1. Regressão Linear Simples
  - a. Objetivo: Avaliar a relação direta entre horas em redes sociais e produtividade
  - b. Métricas:
    - i.  $R^2$
    - ii. MAE
2. Random Forest Regressor
  - a. Objetivo: Capturar relações não-lineares e interações entre variáveis
  - b. Métricas:
    - i.  $R^2$
    - ii. MAE