

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 : 0.45
 - ii. MAE: 12.3 pontos
2. Random Forest Regressor
 - a. Objetivo: Capturar relações não-lineares e interações entre variáveis
 - b. Métricas:
 - i. R^2 : 0.62
 - ii. MAE: 9.8 pontos
 - c. Ajuste de parâmetros:
 - i. Número de estimadores: 100
 - ii. Profundidade máxima: 5
3. Modelo de Baseline (comparação)
 - a. Média simples da produtividade
 - b. MAE: 15.2 pontos

Resultados Obtidos

1. **Relação Negativa:** Foi identificada uma correlação negativa moderada (-0.52) entre tempo em redes sociais e produtividade.

2. **Fatores Adicionais:** O modelo final mostrou que horas de sono e nível de estresse também são preditores importantes da produtividade.
3. **Limiar de Uso:** A análise sugere que até 2 horas diárias em redes sociais têm impacto mínimo, mas além disso a produtividade começa a declinar.
4. **Performance dos Modelos:**
 - a. O Random Forest superou significativamente a regressão linear e o modelo de baseline
 - b. O erro médio absoluto de 9.8 pontos na escala de produtividade (0-100) foi considerado aceitável

Conclusões e Próximos Passos

Os resultados confirmam a hipótese inicial de que o uso excessivo de redes sociais está associado à menor produtividade. No entanto, é importante destacar algumas limitações:

1. **Dados Auto-reportados:** As medidas de produtividade e tempo em redes sociais são subjetivas
2. **Relação de Causalidade:** O estudo identifica correlações, mas não pode estabelecer causalidade

Como próximos passos, sugere-se:

1. Coletar dados objetivos de uso de aplicativos (via APIs)
2. Incluir medidas concretas de produtividade (tarefas concluídas, prazos cumpridos)
3. Desenvolver um modelo preditivo para sugerir limites ideais de uso.