

**Bootcamp Data Analytics 2024**

Desafio Final: Previsão de Bullying



**Desafio Final: Previsão de Bullying**

## **Sobre o conjunto de dados**

## O Global School-Based Student Health Survey (GSHS) é uma pesquisa escolar que usa um questionário autoaplicável para obter dados sobre o comportamento de saúde dos jovens e fatores de proteção relacionados às principais causas de morbidade e mortalidade. A pesquisa foi realizada na Argentina em 2018. Um total de 56.981 alunos participaram. A taxa de resposta da escola foi de 86%, a taxa de resposta do aluno foi de 74% e a taxa de resposta geral foi de 63%.

**Sobre o desenvolvimento de um modelo de previsão de bullying**

A partir do GHSH, foram selecionadas as perguntas da pesquisa relacionadas ao bullying. Isto com o objetivo de desenvolver um modelo que permita preveni-la.  
Alguns dos indicadores de bullying têm sido relacionados a sentimentos de solidão, falta de amigos íntimos, má comunicação com os pais, ausência nas aulas. (por exemplo, Nansel et al. Comportamentos de bullying entre jovens americanos: Prevalência e associação com ajustamento psicossocial)  
Da mesma forma, jovens identificados por agressores como abaixo do peso, acima do peso e obesos foram vítimas de bullying.

Para responder esse desafio: A base de dados contém as seguintes informações:

* record : código identificador
* Bullied\_on\_school\_property\_in\_past\_12\_months: Sofreu bullying na propriedade da escola nos últimos 12 meses
* Bullied\_not\_on\_school\_property\_in\_past\_12\_months: sofreu bullying fora da propriedade da escola nos últimos 12\_meses
* Cyber\_bullied\_in\_past\_12\_months: Cyberbullying nos últimos 12 meses
* Custom\_Age: Idade
* Sex: Sexo
* Physically\_attacked: Ataque físico
* Physical\_fighting: Briga Física
* Felt\_lonely: Sentiu-se sozinho
* Close\_friends: Amigos próximos
* Miss\_school\_no\_permission: Faltam à escola sem permissão
* Other\_students\_kind\_and\_helpful: Outros alunos gentis e prestativos
* Parents\_understand\_problems: Os pais entendem os problemas
* Most\_of\_the\_time\_or\_always\_felt\_lonely: Na maioria das vezes ou sempre se sentiu sozinho
* Missed\_classes\_or\_school\_without\_permission: Perdeu aulas ou escola sem permissão
* Were\_underweight: Estava abaixo do peso
* Were\_overweight: Estava acima do peso
* Were\_obese: Estava obeso

1. **Análise dos dados:**
   * Calcule a matriz de correlação para todas as variáveis numéricas e Identifique as correlações mais fortes e mais fracas
   * Faça a análise exploratória dos dados categóricos, calcule a moda e as variáveis e utilize recursos visuais, utilize o teste qui-quadrado para avaliar a análise de correspondência.
   * Plote histogramas e boxplots para cada tipo de bullying separado por gênero e idade.

### **2. Modelo de Regressão Logística**

* + Construa um modelo de regressão logística para prever a probabilidade de um indivíduo sofrer bullying na escola, considere como target: Bullied\_on\_school\_property\_in\_past\_12\_months que representa 1 caso o indivíduo tenha sofrido bullying na escola nos últimos 12 meses, 0 caso contrário.
  + Análise as variáveis mais relevantes no modelo.
  + Avalie a performance do modelo usando métricas como Acurácia, Precisão, Recall e plote a matriz de confusão.