

# Desafio Celero - Cientista de dados.

## Objetivo

A finalidade deste desafio é testar as suas habilidades e conhecimento de algoritmos de classificação com base em aprendizado de máquinas supervisionado, assim como a implementação destas soluções utilizando a linguagem de programação Python.

O seu objetivo será criar classificador de comentários de filmes. A base de dados para treino pode ser acessada em <http://ai.stanford.edu/~amaas/data/sentiment/>.

- Quaisquer bibliotecas Python podem ser usadas para solucionar o problema.
- O programa deve possuir duas configurações separadas: uma para treino e outra para execução do algoritmo usando o modelo treinado.
  - Quando estiver em modo de treinamento o programa deve receber como parâmetro um diretório com os comentários em inglês dentro de arquivos texto e o resultado daqueles comentários (1 para comentários positivos e -1 para comentários negativos).
  - Quando estiver em modo de execução deve receber como parâmetro um arquivo texto contendo um comentário em inglês e deve retornar o número 1 para comentários positivos e -1 para comentários negativos.
- Você deverá escrever um breve relatório para explicar os conceitos usados e quais os resultados obtidos, como comparação entre os modelos usados e qual modelo teve uma melhor performance e o porquê.
- Toda a informação para rodar e usar o projeto deve estar no arquivo README.md.

## PONTOS BÔNUS

- Legibilidade e documentação do código;
- Performance do algoritmo (Tanto tempo de execução quanto acerto).

## Prazo

Você terá 1 (uma) semana para concluir o desafio.

## Avaliação

- A utilização das tecnologias requeridas;
- Se o seu resultado possui um bom nível de confiabilidade;
- A forma com a qual você escolheu escrever o desafio (a arquitetura que usou, se está bem documentado, se possui uma boa legibilidade, etc);
- Como escolheu organizar o seu fluxo de trabalho no git (como uso de branches, separação das responsabilidades, etc).
- Se o seu teste é genuíno, ou seja, se foi você quem o escreveu (evite soluções prontas, isto causará a sua imediata desclassificação).
- A elaboração do relatório para descrever o seu resultado.

## Entrega

Você deve versionar o seu desafio no **git** desde o início. Para hospedá-lo remotamente, escolha entre **Github** e **Bitbucket**. Depois de empurrar o seu progresso à plataforma escolhida, compartilhe conosco o link do mesmo.