



# Competição do Kaggle - 2

## Aprendizado de Máquina

Prof. Me. Otávio Parraga

---

## Introdução

A plataforma [Kaggle](#) é um ambiente online que reúne cientistas de dados para resolver problemas reais por meio de competições. Nesta atividade, você utilizará ferramentas como [Pandas](#) (para manipulação de dados) e [Scikit-Learn](#) (para modelos de machine learning) para desenvolver um classificador para analisar sentimentos. Ambientes interativos como [Jupyter Notebooks](#), [Google Colab](#) ou Kaggle Notebooks são recomendados para implementação e teste do código.

---

## Objetivo e Tarefas

O objetivo da atividade é desenvolver um modelo que consiga classificar se um determinado comentário sobre uma revista (review) é negativo (bad) ou positivo (good). Geralmente, quando temos dados textuais ligados a um sentimento, tratamos esse problema como sendo uma tarefa específica de classificação chamada de Análise de Sentimentos. Para essa finalidade, utilize da biblioteca [Scikit-Learn](#) para instanciar os modelos vistos em aula ( [K-Nearest Neighbors \(KNN\)](#), [Naïve Bayes](#) e [Árvores de Decisão](#)) e da biblioteca Pandas para ler os dados como uma tabela.

O conjunto de dados disponibilizado possui três arquivos: um contendo os dados de treinamento (anotados), um como o conjunto de teste (sem as anotações) e um arquivo de exemplo para submissão. Os testes e modelagem devem ser realizados utilizando o conjunto de treinamento; o conjunto de teste serve apenas para gerar o arquivo de submissão. Ao gerar o arquivo, você deve fazer a submissão do mesmo no link da competição. Os resultados serão avaliados automaticamente e um valor de escore será atribuído.

A atividade pode ser feita individualmente ou em grupos de até 3 pessoas. Utilize os notebooks de exemplo disponíveis no Moodle como referência.

---

## Pontos Extra (0.5 na Primeira Prova)

Para ganhar a bonificação, é necessário:

- Atingir acurácia mínima de 87% no desafio.
- Postar o notebook completo que gerou o resultado no fórum do Moodle, garantindo que os resultados sejam reproduzíveis.
- Apenas um integrante do grupo deve enviar a submissão final no Kaggle.