

Fresco ou Podre?

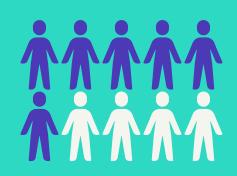
uma análise dos reviews do Rotten Tomatoes

Analisar *reviews* de críticos de filmes e aplicar sobre reviews de usuários para obter uma classificação destes.

DE ONDE VIERAM OS DADOS?

Kaggle: Dados dos filmes e dos reviews dos usuários.

Rotten Tomatoes: dados do filme a ser analisado.



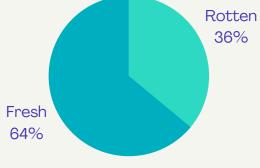
FORMATO DOS DADOS

Dados de *reviews* dos críticos: texto em linguagem natural, majoritariamente em língua inglesa.

Rótulo dos reviews: "Rotten" ou "Fresh".

Dados dos *reviews* dos usuários: texto em linguagem natural em diversas línguas. Dados sem rótulos.





Distribuição dos rótulos nos dados de reviews dos críticos.



METODO LOGIAS

Aprendizado de máquina supervisionado TF-IDF Classificadores

Utilizar os textos em linguagem natural para treinar os modelos. Lematizar e manter apenas os textos em inglês para treinar os modelos.



RESULTADOS

Um modelo treinado capaz de, dado um texto de *review* qualquer, classificá-lo como *Rotten* ou *Fresh*.



Reviews de usuários classificados como Rotten



Acurácia*

Área sob
a curva*

80%

* da Regressão Logística, o modelo escolhido