

## Exercício Final – Projeto de Análise de Sentimento em Literatura Vitoriana

 Tema: "Emoções na Literatura Clássica Inglesa"

Neste desafio final, vais aplicar os conhecimentos adquiridos ao longo do workshop para construir um modelo de análise de sentimento sobre frases extraídas de obras literárias da era vitoriana (século XIX). O objetivo será **classificar automaticamente o tom emocional** de excertos literários, usando abordagens supervisionadas de NLP.

---

### O Dataset: VictorianLit

 Link para o dataset:

 <https://www.kaggle.com/datasets/elibooklover/victorianlit>

 Página com explicação do projeto:

 <https://elibooklover.github.io/VictorianLit>

Este dataset contém **53 805 frases** extraídas de obras clássicas da literatura vitoriana, como os romances de *Jane Austen*, *Charles Dickens*, *Emily Brontë* e outros autores influentes da época.

Cada frase está anotada com um **nível de sentimento** que varia de 0 a 4:

- 0 – Muito Negativo (*Very Negative*)
  - 1 – Negativo (*Negative*)
  - 2 – Neutro (*Neutral*)
  - 3 – Positivo (*Positive*)
  - 4 – Muito Positivo (*Very Positive*)
- 

### Transformação para Classificação Binária

Para simplificar a tarefa, vamos transformar esta classificação em apenas **duas classes (binária)**:

Original	Reclassificado	Interpretação
0	0	Negativo
1	0	Negativo
2	✗ Remover	Neutro — será excluído
3	1	Positivo
4	1	Positivo

 Ou seja, vamos apenas considerar frases **claramente negativas ou positivas**, descartando as neutras (2).

---

### Objetivo do Projeto

Construir um modelo capaz de **classificar frases literárias como positivas ou negativas**.

Para isso, deverá:

- Efetuar uma **limpeza e pré-processamento do texto** (remoção de emojis, stopwords, negação, etc.).
- Representar os textos numericamente (TF-IDF, embeddings, etc.).
- Treinar **um ou mais modelos de machine learning** (por exemplo: regressão logística, SVM, redes neurais).
- Avaliar o desempenho com métricas apropriadas (acurácia, F1, matriz de confusão).
- (Opcional) Visualizar os resultados e padrões no texto.