Coleções de dados

* <https://github.com/franciellevargas/HateBR>
* <https://huggingface.co/datasets/dougtrajano/olid-br> / <https://www.kaggle.com/datasets/dougtrajano/olidbr> /
* <https://github.com/JAugusto97/ToLD-Br>

1. Abordagem : Um Trabalho em que se use uma coleção de dados única unindo todas as coleções de dados
2. Abordagem : Um trabalho que analise-se as coleções de dados distintas utilizando um único classificador

* Na etapa de pré-processamento, tomar cuidado com os sinais/emojis que possam ser direcionados para algum grupo marginalizado específico
* Pensar na possibilidade de incluir esses emojis, caso não haja.

Tutorial : <https://github.com/jiayao99/tfidf-text-classification/blob/main/resume_classification_tfidf.ipynb>