

Desafio

Você deverá utilizar os dados para criar um modelo que classifique discurso de ódio no dataset fornecido.

O resultado do trabalho deverá estar em jupyter notebook com as seguintes tópicos:

- Análise exploratório dos dados (EDA).
- Pré-processamento.
- Escolha das features.
- Criação do modelo de Machine Learning (não use deep learning para essa tarefa).
- Treinamento.
- Escolha da métrica (Recall, F1, acurácia ou precisão).
- Avaliação dos resultados de forma visual através da métrica escolhida.

Observações

- Coloque o jupyter notebook em uma imagem docker e suba ela no dockerhub. Responda o e-mail com o link para esse registry. A avaliação do notebook será exclusivamente via este registry.
- Utilize python versão 3.
- Libere o código no github (jupyter, arquivo docker, etc).
- Use exclusivamente o dataset de treino para esta tarefa.
- Tente justificar ao máximo as suas decisões dentro do jupyter notebook.

O dataset esta disponível em:

<https://www.kaggle.com/arkhoshghalb/twitter-sentiment-analysis-hatred-speech#train.csv>