

Nome: Morsinaldo de Azevedo Medeiros
Matrícula: 20220044322
Curso: Engenharia de Computação

Olá, pessoal! Minha intenção inicial era criar um modelo utilizando redes neurais para prever o assunto processual a partir das palavras da coluna *conteudo_sentenca*, mas acabou que a RAM do google colab sempre estourava e reiniciava a seção. Eu até consegui treinar a rede em outro notebook como prova de conceito, mas quando adaptei para o pipeline deu esse problema. Esse notebook está com o nome *end-to-end* na pasta chamada *prova de conceito*, mas é apenas para o caso de vocês quiserem analisar mesmo. Os notebooks principais são os da pasta raiz (1 ao 7). Eu coloquei os artefatos gerados na pasta *Artefatos*, mas todos se encontram no wandb também (link abaixo), eu deixei o projeto público para vocês verem. Caso vocês queiram gerar os mesmos artefatos, é só fazer o login no wandb.

Devido ao problema da rede neural, eu não consegui fazer o deploy no Heroku como eu havia planejado inicialmente, mas me empenhei muito e acredito que fiz o máximo possível, no tempo que eu tive para estudar sobre processamento de linguagem natural. No final, eu acabei utilizando uma biblioteca chamada BerTopic para fazer modelagem baseada em tópicos, a qual eu não conhecia também.

Eu tinha uma noção de NLP, mas nunca peguei um problema para realizar essas análises, então, foi de grande aprendizado para mim. Tanto nessa parte, quanto a parte jurídica, visto haver muitos conceitos desconhecidos para mim, até então. Espero nos encontrarmos na fase da entrevista e espero que vocês gostem do que eu fiz também. Muito obrigado pela oportunidade!

OBS.: Não consegui enviar junto o arquivo do modelo por que ficou com 450 mb e excedia o tamanho do arquivo para enviar, mas todos os artefatos podem ser baixados através do wandb.

link para o wandb: https://wandb.ai/morsinaldo/nlp_bolsa?workspace=user-morsinaldo
link para o github: https://github.com/Morsinaldo/Selecao_NLP