

Trabalho 4 - Classificando notícias

Vamos classificar notícias como verdadeiras ou falsas?

VISÃO GERAL

Nesse exercício construiremos utilizaremos PLN para transformar nosso conjunto de dados e classificá-lo como notícias verdadeiras ou falsas.

OBJETIVOS

1. Representar um conjunto de dados real de textos de maneira vetorial;
2. Comparar as técnicas de vetorização Count Vectorizer e TF-IDF Vectorizer;
3. Classificar textos com o algoritmo Naive Bayes;
4. Comparar os resultados da classificação utilizando precisão e matriz de confusão.

ESPECIFICAÇÕES

Neste trabalho, você deverá seguir os seguintes passos

1. Importar e separar o conjunto de dados fornecido em treinamento e teste;
2. Criar uma matriz vetorizada que represente os textos utilizando o `CountVectorizer`;
3. Criar uma matriz vetorizada que represente os textos utilizando o `TfidfVectorizer`;
4. Criar dois modelos para classificar os textos utilizando o algoritmo Naive Bayes (multinomial) utilizando o `CountVectorizer` e o `TfidfVectorizer`;
5. Comparar os resultados utilizando a precisão e a Matriz de Confusão;
6. [BÔNUS] Propor passos adicionais para melhorar os resultados encontrados.

AVALIAÇÃO

Você deverá enviar o link do notebook no Google Colab. Cada passo listado nas especificações do trabalho será avaliado.