

Oficina_PLN

June 30, 2022

0.0.1 Bibliotecas necessárias

```
[ ]: ! pip install nltk
```

```
Requirement already satisfied: nltk in /home/rafael/Learn/Python/Introdução a
Machine Learning/env/lib/python3.10/site-packages (3.7)
Requirement already satisfied: regex>=2021.8.3 in
/home/rafael/Learn/Python/Introdução a Machine Learning/env/lib/python3.10/site-
packages (from nltk) (2022.6.2)
Requirement already satisfied: tqdm in /home/rafael/Learn/Python/Introdução a
Machine Learning/env/lib/python3.10/site-packages (from nltk) (4.64.0)
Requirement already satisfied: click in /home/rafael/Learn/Python/Introdução a
Machine Learning/env/lib/python3.10/site-packages (from nltk) (8.1.3)
Requirement already satisfied: joblib in /home/rafael/Learn/Python/Introdução a
Machine Learning/env/lib/python3.10/site-packages (from nltk) (1.1.0)
```

```
[ ]: import nltk
from nltk.tokenize import word_tokenize
from sklearn.feature_extraction.text import TfidfVectorizer
import pandas as pd
```

0.0.2 Pacotes necessários

```
[ ]: nltk.download('stopwords')
stopwords = nltk.corpus.stopwords.words('portuguese')
nltk.download('punkt')
```

```
[nltk_data] Downloading package stopwords to /home/rafael/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to /home/rafael/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

```
[ ]: True
```

Use a variável texto a seguir para identificar as stop words e após isso aplique o cálculo do **TF-IDF** neste mesmo texto e observe os scores.

OBS: Para realizar essa tarefa utilize os módulos estudados na aula

```
[ ]: texto = 'Um carro azul seguia rapidamente em uma rodovia, e ao passar por um
↳buraco, o carro furou o pneu, e o motorista desceu do carro azul'
```

```
[ ]: tf_idf = TfidfVectorizer()
vetor = tf_idf.fit_transform([texto])
vetor = vetor.todense()
nomes = tf_idf.get_feature_names_out()
```

```
[ ]: tokens = word_tokenize(texto)

for t in tokens:
    if t not in stopwords:
        print(t)
```

```
Um
carro
azul
seguia
rapidamente
rodovia
,
passar
buraco
,
carro
furou
pneu
,
motorista
desceu
carro
azul
```

```
[ ]: df = pd.DataFrame(vetor, columns=nomes)
print(df)
```

```

      ao      azul      buraco      carro      desceu      do      em \
0  0.179605  0.359211  0.179605  0.538816  0.179605  0.179605  0.179605

      furou  motorista      passar      pneu      por  rapidamente  rodovia \
0  0.179605  0.179605  0.179605  0.179605  0.179605  0.179605  0.179605

      seguia      um      uma
0  0.179605  0.359211  0.179605
```