

Disciplina: Machine Learning I
Professora: Cristiane Neri Nobre
Data de entrega: 09/06
Valor: 4 pontos

Questão 01

No endereço

<https://www.kaggle.com/code/leandrodoze/sentiment-analysis-in-portuguese/notebook>

Você terá acesso a um código para análise de sentimentos usando o conceito de reconhecimento de texto.

Veja o conceito de Reconhecimento em Entidades Nomeadas

<https://drive.google.com/drive/folders/1eBdTcmBA5Z7ALFEtfBEyBNAYNbYsJifa>

Veja também: https://www.tensorflow.org/text/guide/word_embeddings

O que você precisa fazer:

- 1) Experimentar este código e se basear nele para resolver o problema da base de dados disponível no CANVAS: **ReutersGrain-train.csv** e **ReutersGrain-test.csv**

Esta base de dados é de classificação e classifica se o texto refere-se a **grãos ou não**.

Desta forma, após a preparação do texto, você deverá utilizar qualquer algoritmo de aprendizado (Random Forest, Árvore de Decisão, Naive, etc) para resolver o problema.

Algumas importações importantes que vocês irão precisar:

```
import string
import pandas as pd
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn import metrics
from sklearn.model_selection import cross_val_predict
```

```
import seaborn as sns
from matplotlib import pyplot as plt
from sklearn.model_selection import GridSearchCV
```

```
nltk.download('punkt')
nltk.download('stopwords')
stemmer = PorterStemmer()
```