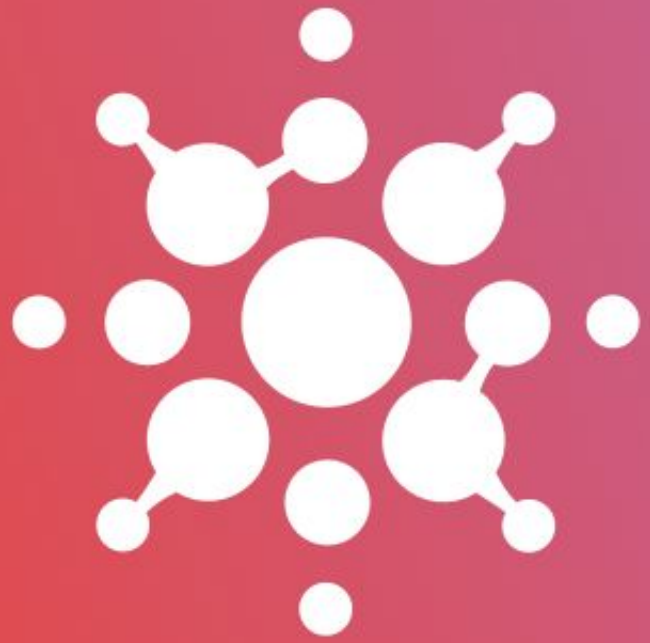
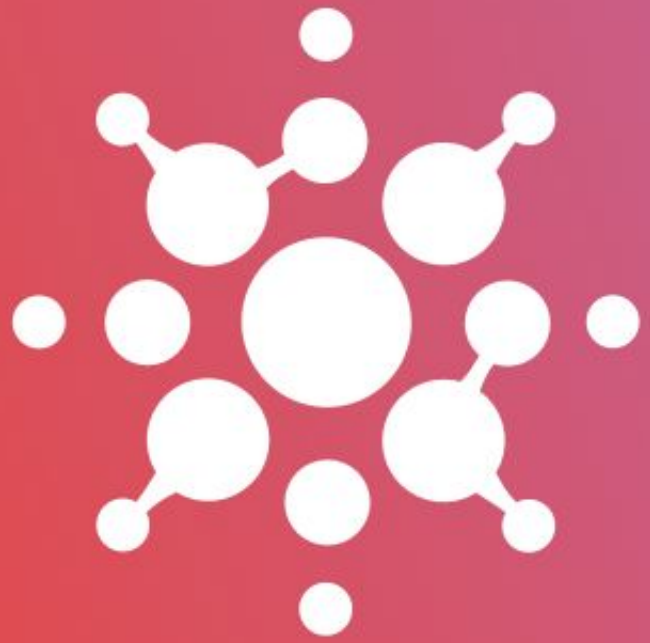


AI Brasil

an artificial intelligence community



AI Brasil

an artificial intelligence community

Rotulando e treinando um Classificador para NLP

Meetup AI Brasil – 23/08/19 @ Hotmilk

CARA

...

- Maratonista de séries
- Gosta de ajudar
- Programa desde os 15 anos

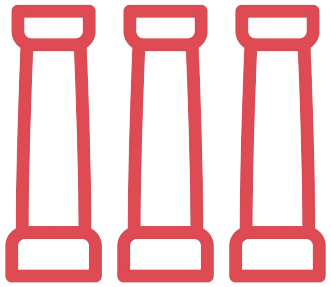
... E CRACHÁ!

- 10 anos de carreira sendo 4 em ML
- ML Developer @ PhoneTrack
- Bacharel em Sistemas de Informação

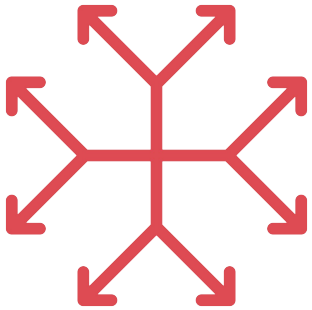
Linkedin> fillipedornelas
Instagram> @fdornelasx

AI Brasil
an artificial intelligence community

Agenda



Princípio



Possibilidades

DEMO &
CODE

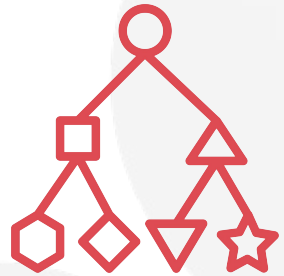
Fundamentos

AI Brasil
an artificial intelligence community

O Processamento de Linguagem Natural (PLN) é a subárea da Inteligência Artificial (IA) que estuda a capacidade e as limitações de uma máquina em entender a linguagem dos seres humanos.

ADQUIRIR DADOS

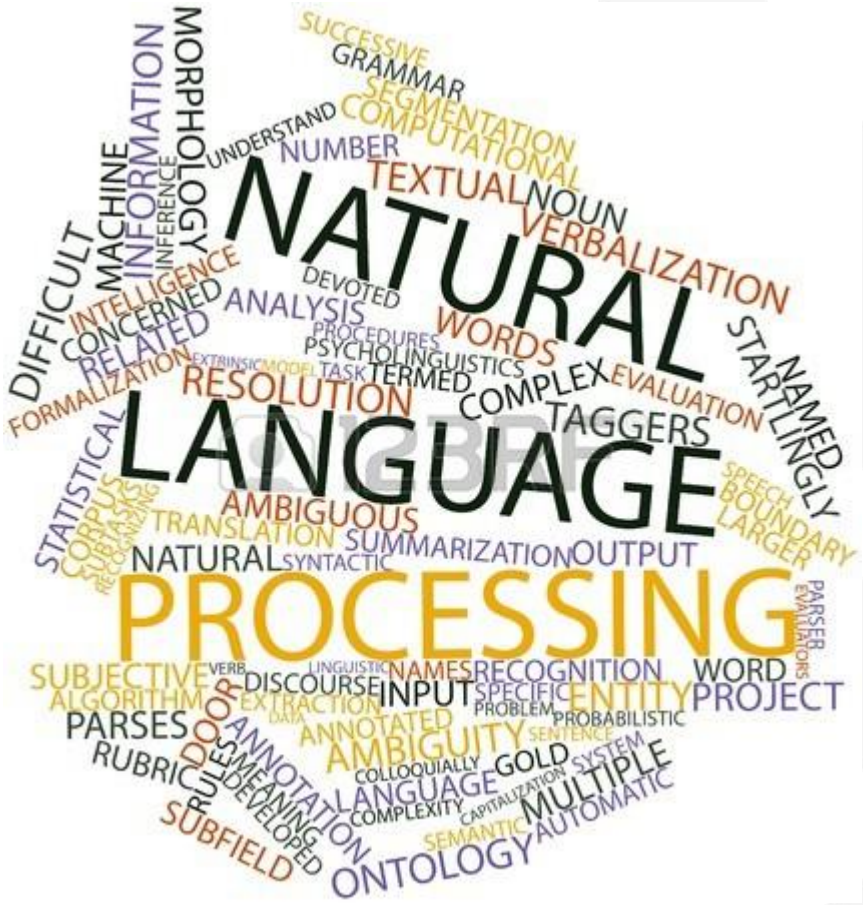
ROTULAR



CLASSIFICAR

Possibilidades

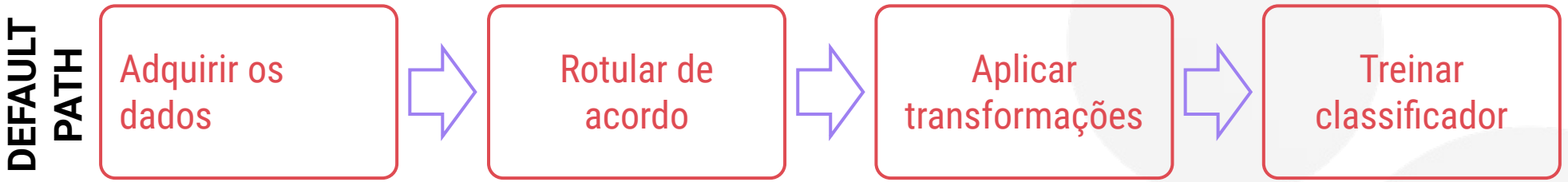
AI Brasil
an artificial intelligence community



Etapas

Pipeline Funcional

AI Brasil
an artificial intelligence community



NLP - Pipeline

Adquirir e Rotular

AI Brasil
an artificial intelligence community

Reviews do IMDB (+49k reviews)	**Classes**
• Mais uma vez, o Sr. Costner arrumou um filme por muito mais tempo do que o necessário.	Negativo
• Isto é verdadeiramente, sem exagerar, um dos piores filmes de Slasher já feitos.	Negativo
• Este é o tipo de filme que você quer ver com um copo de vinho, o fogo aceso e com os pés para cima.	Positivo
• Eu vi o filme recentemente e realmente gostei. Eu me surpreendi e chorei.	Positivo

NLP - Vetorizando

Codificar o texto

AI Brasil
an artificial intelligence community

Reviews do IMDB (+49k reviews)

- Mais uma vez, o Sr. Costner arrumou um filme por muito mais tempo do que o necessário.
- Isto é verdadeiramente, sem exagerar, um dos piores filmes de Slasher já feitos.
- Este é o tipo de filme que você quer ver com um copo de vinho, o fogo aceso e com os pés para cima.
- Eu vi o filme recentemente e realmente gostei. Eu me surpreendi e chorei.

Scikit Learn - CountVectorizer

- 47843 76531 77703 24 54062 70837 89 18961 6917 76526 32825 58919 51992 47843 73515 24831 61665 54062 52895 89
- 42752 80491 67988 76526 25239 58207 32890 20699 69654 43892
- 29858 80491 54062 74303 20699 32825 61665 78444 61760 77483 16304 76526 18545 20699 54062 33545 2159 25931 16304 55299 61434 56021
- 30396 77723 54062 32825 62875 25931 62682 30396 49476 72363 25931

NLP - TF-IDF

Codificar face

AI Brasil
an artificial intelligence community

Scikit Learn - CountVectorizer

- 47843 76531 77703 24 54062 70837 89 18961 6917 76526 32825 58919 51992 47843 73515 24831 61665 54062 52895 89
- 42752 80491 67988 76526 25239 58207 32890 20699 69654 43892
- 29858 80491 54062 74303 20699 32825 61665 78444 61760 77483 16304 76526 18545 20699 54062 33545 2159 25931 16304 55299 61434 56021
- 30396 77723 54062 32825 62875 25931 62682 30396 49476 72363 25931

Scikit Learn - TfidfTransformer

- 0.05859714431066524 0.08704830258140413 0.038583589132334915 0.13388660619947934 0.03777333617823882 0.050793874567107156 0.03848636391503398 0.07846683589457548 0.037852511917520794 0.032140229312573 55

- **Enfim... acho que vocês entenderam ;P**

Treino

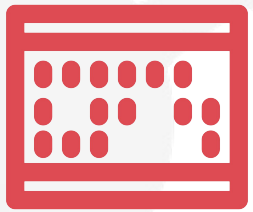
Processo completo até o treino

AI Brasil
an artificial intelligence community

Mais uma vez, o Sr. Costner arrumou um filme por muito mais tempo do que o necessário.

42752 80491
67988 76526
25239 58207
32890 20699
69654 43892

0.05859714431066524
0.08704830258140413
0.038583589132334915
0.13388660619947934
0.03777333617823882
0.050793874567107156
0.03848636391503398
0.07846683589457548
0.037852511917520794
0.03214022931257355



Texto Rotulado

Vectorizer

TFIDF

Treino

Treino - Final

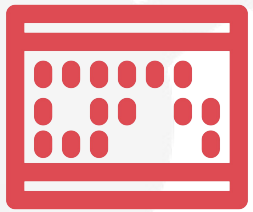
Processo completo até o resultado

AI Brasil
an artificial intelligence community

Mais uma vez, o Sr. Costner arrumou um filme por muito mais tempo do que o necessário.

42752 80491
67988 76526
25239 58207
32890 20699
69654 43892

0.05859714431066524
0.08704830258140413
0.038583589132334915
0.13388660619947934
0.03777333617823882
0.050793874567107156
0.03848636391503398
0.07846683589457548
0.037852511917520794
0.03214022931257355



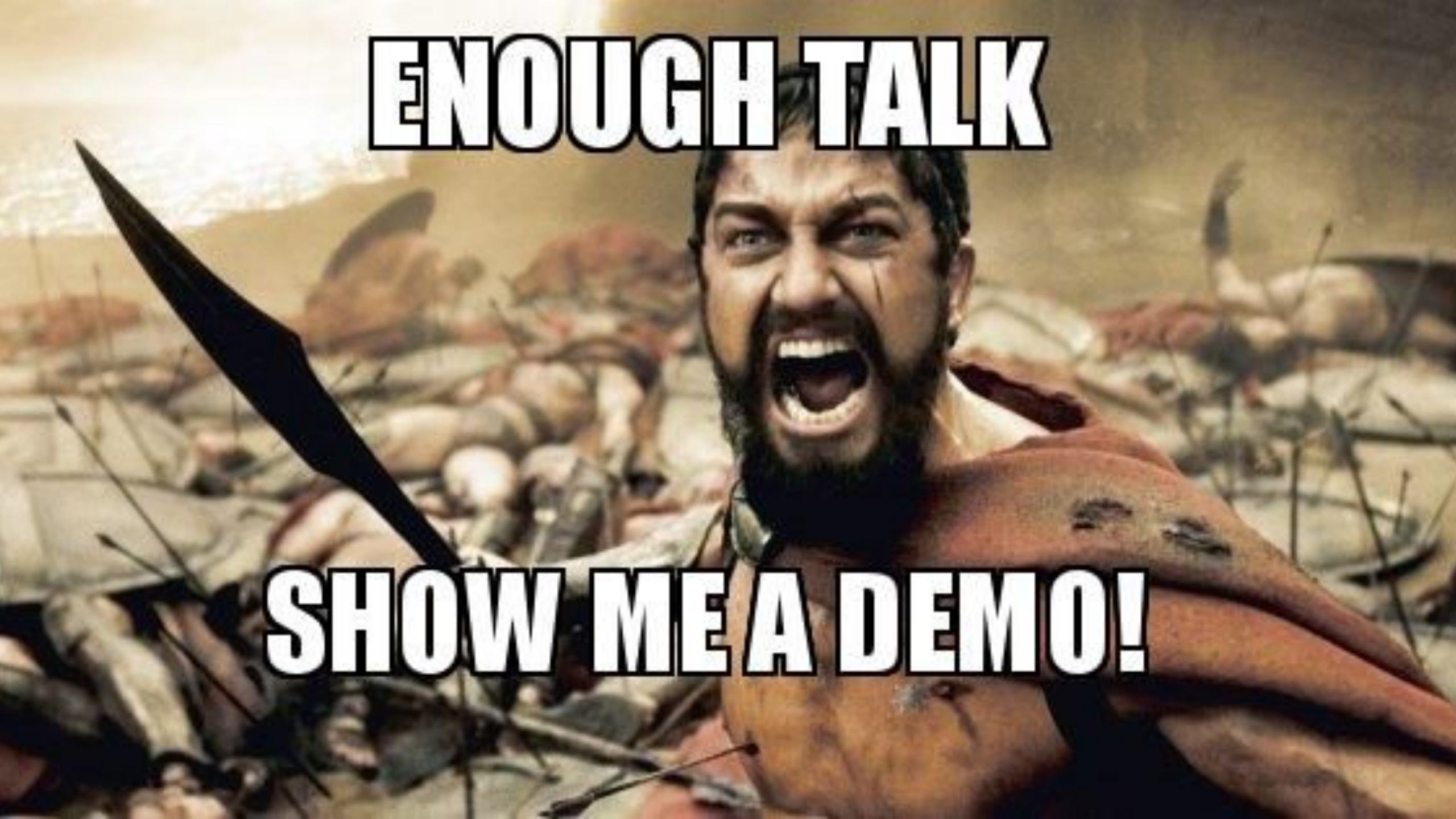
Negativo

Texto Rotulado

Vectorizer

TFIDF

Treino


ENOUGH TALK
SHOW ME A DEMO!

AI Brasil

an artificial intelligence community

Dúvidas?