

Processamento de Linguagem Natural - 2022.1

Apresentação Final

Álvaro Amorim - aaa2@ic.ufal.br
Rafael Laranjeira - rml@ic.ufal.br

Conjunto de Dados

Practical solutions to the problem of diagonal dominance in kernel document clustering, Proc. ICML 2006.

Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering

Derek Greene
Pádraig Cunningham

University of Dublin, Trinity College, Dublin 2, Ireland

DEREK.GREENE@CS.TCD.IE
PADRAIG.CUNNINGHAM@CS.TCD.IE

Abstract

In supervised kernel methods, it has been observed that the performance of the SVM classifier is poor in cases where the diagonal entries of the Gram matrix are large relative to the off-diagonal entries. This problem, referred to as *diagonal dominance*, often occurs when certain kernel functions are an-

can limit the extent to which the solution space is explored beyond the initial state.

An unfortunate characteristic of this problem is that matrices which are strongly diagonally dominated will be positive semi-definite and measures to reduce this dominance run the risk of rendering the matrix indefinite so that it no longer represents a valid Mercer kernel. Consequently, there is a tension between

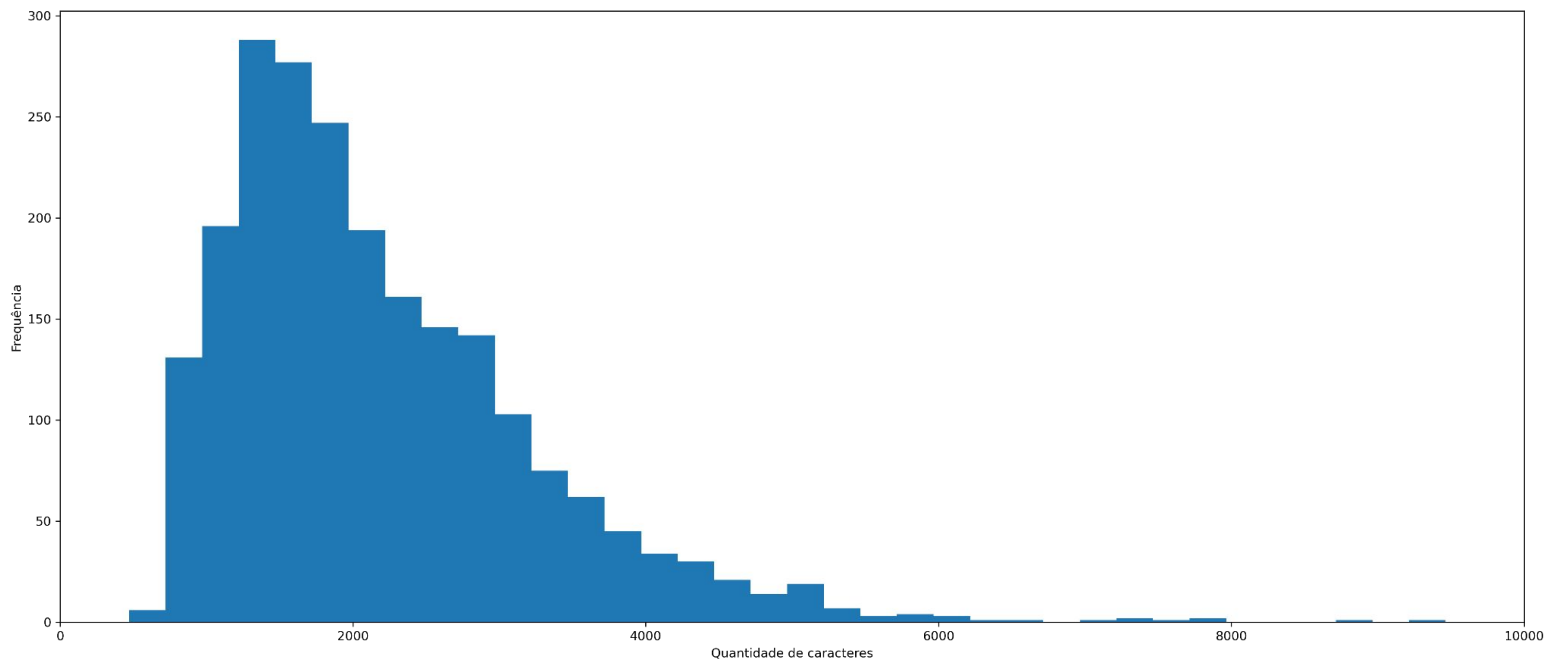
Classe	Quantidade
business	510
entertainment	386
politics	417
sport	511
tech	401
Total	2225

Derek Greene and Pádraig Cunningham. 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proceedings of the 23rd international conference on Machine learning (ICML '06)*. Association for Computing Machinery, New York, NY, USA, 377–384. <https://doi.org/10.1145/1143844.1143892>

Lista 1

Questão 06

Determine a distribuição de comprimentos dos textos (em quantidade de caracteres), listando estas quantidades e plotando um histograma.



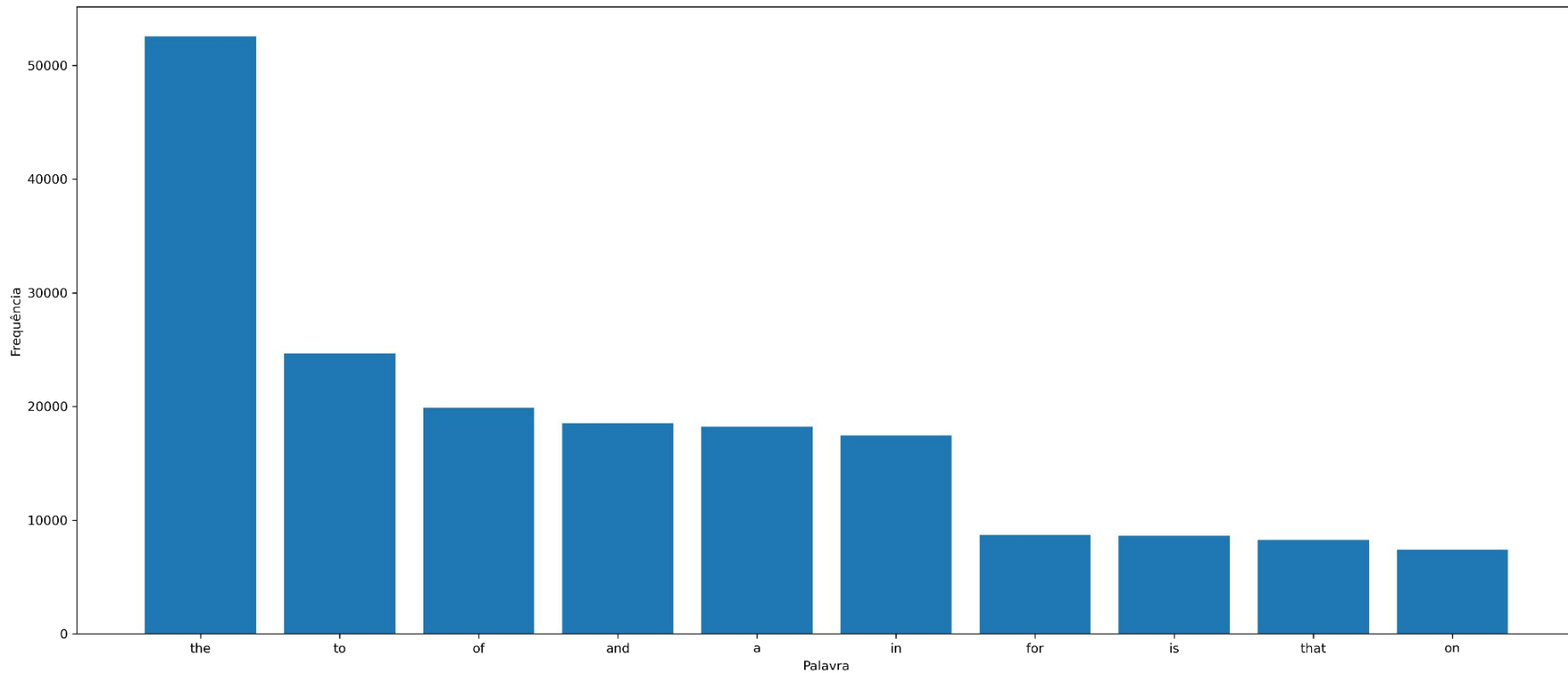
Questão 07

Aplique os seguintes passos de pré-processamento aos textos:

- Remova todas as palavras que contêm números;
- Converta as palavras para minúsculas;
- Remova pontuação;
- Tokenize os textos em palavras, gerando um dicionário único com n tokens e convertendo cada texto em um vetor de dimensão n com a respectiva contagem de palavras.

Em seguida, encontre as 10 palavras mais frequentes da base de textos.

Questão 07



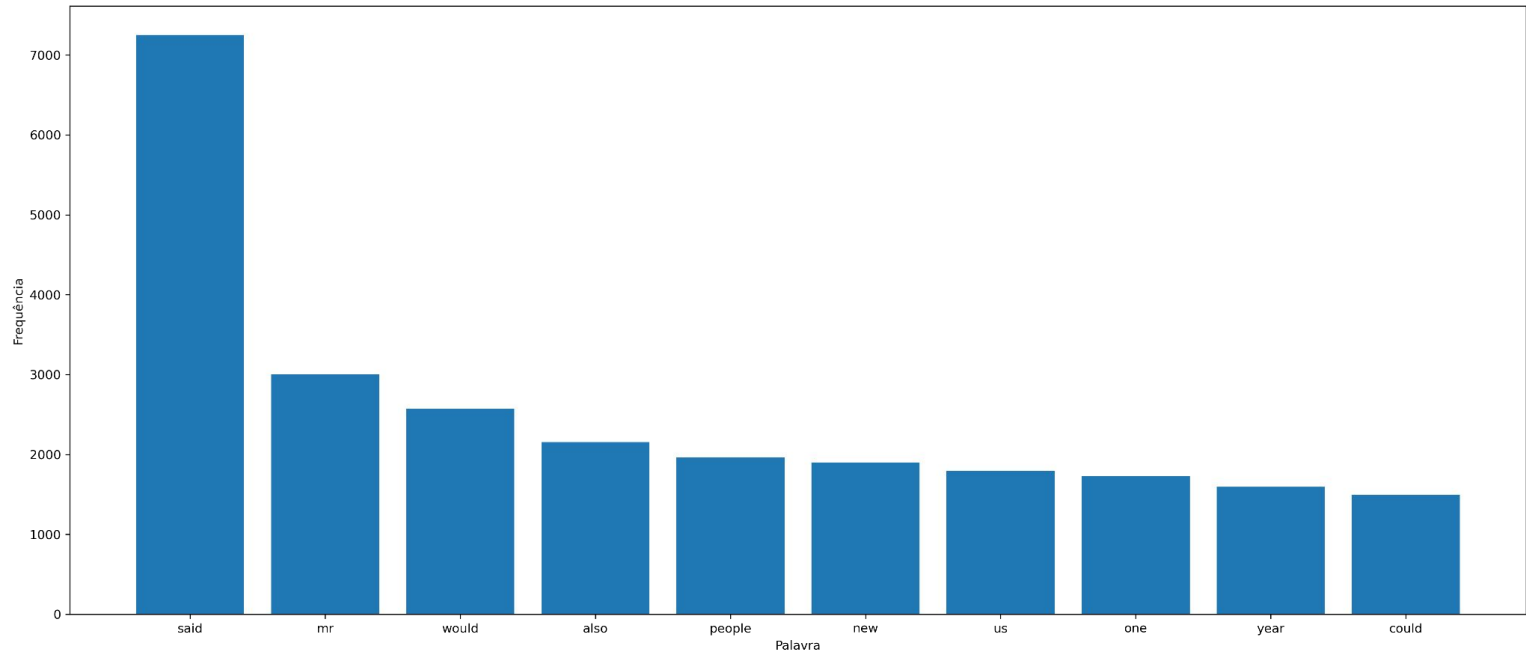
Questão 08

Aplique os seguintes passos de pré-processamento aos textos processados na questão anterior:

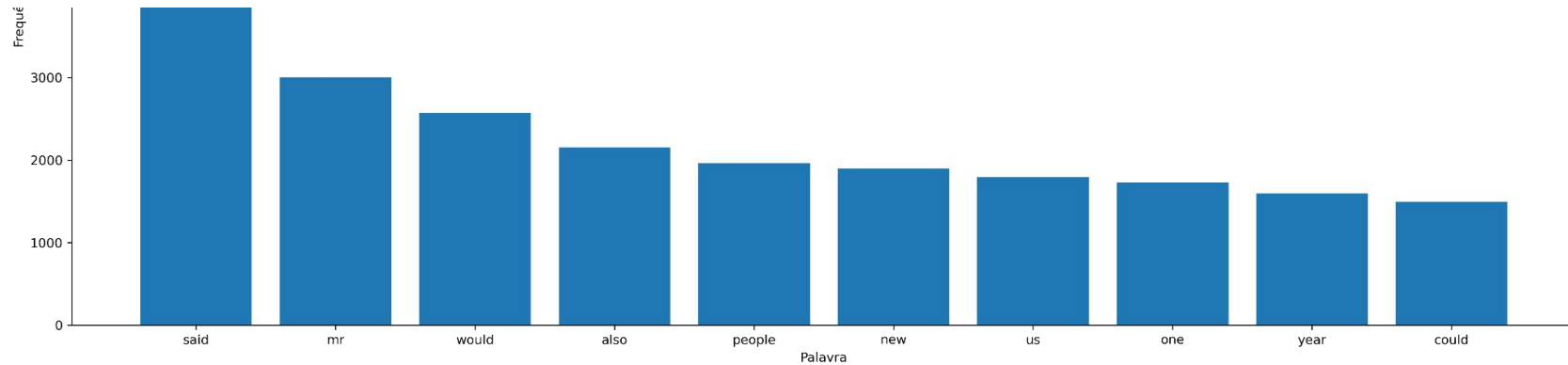
- Remova stopwords;
- Realize rotulação de POS;
- Realize stemização;

Questão 08

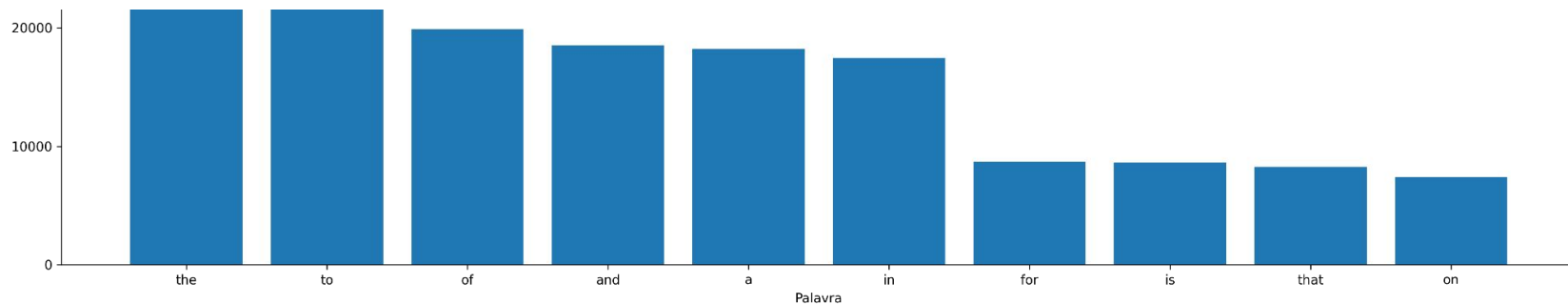
b) Verifique quais são as 10 palavras mais frequentes e compare com as 10 palavras mais frequentes da questão anterior.



Questão 08

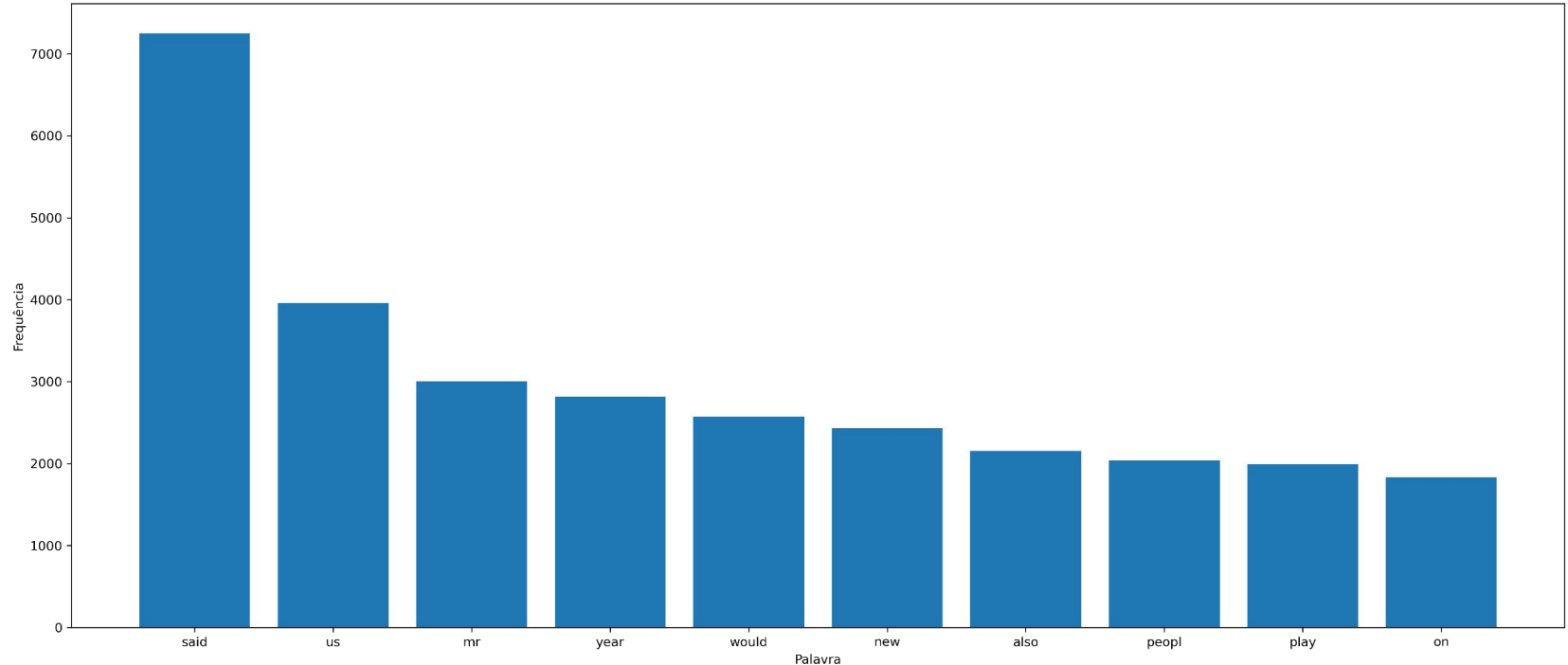


Questão 07



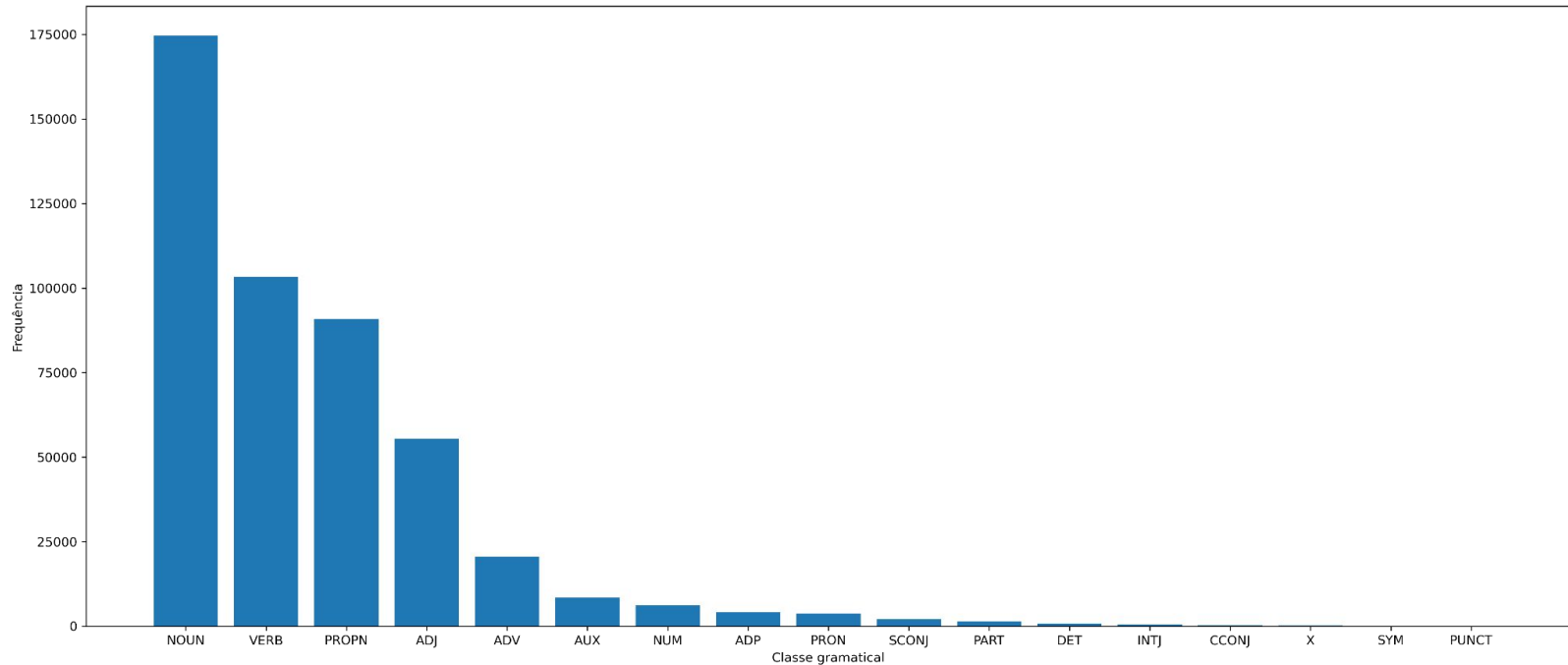
Questão 08

c) Repita a letra b) usando os tokens stemizados.



Questão 08

d) Verifique quais são as classes gramaticais mais frequentes.



Lista 2

Questão 02

Dataset:

- Classe sport virou 1 e as outras viraram 0.

Classe	Quantidade
0	1714
1	511
Total	2225

Questão 02

- Logistic Regression

TF-IDF

		Previu	
		0	1
Verdadeiro	0	515	0
	1	22	131

	precision	recall	f1-score	support
0	0.96	1.00	0.98	515
1	1.00	0.86	0.92	153
accuracy			0.97	668
macro avg	0.98	0.93	0.95	668
weighted avg	0.97	0.97	0.97	668

CountVectorizer

		Previu	
		0	1
Verdadeiro	0	515	0
	1	1	152

	precision	recall	f1-score	support
0	1.00	1.00	1.00	515
1	1.00	0.99	1.00	153
accuracy			1.00	668
macro avg	1.00	1.00	1.00	668
weighted avg	1.00	1.00	1.00	668

Questão 02

- Naive-Bayes

TF-IDF

		Previu	
		0	1
Verdadeiro	0	515	0
	1	16	137

	precision	recall	f1-score	support
0	0.97	1.00	0.98	515
1	1.00	0.90	0.94	153
accuracy			0.98	668
macro avg	0.98	0.95	0.96	668
weighted avg	0.98	0.98	0.98	668

CountVectorizer

		Previu	
		0	1
Verdadeiro	0	512	3
	1	13	140

	precision	recall	f1-score	support
0	0.98	0.99	0.98	515
1	0.98	0.92	0.95	153
accuracy			0.98	668
macro avg	0.98	0.95	0.97	668
weighted avg	0.98	0.98	0.98	668

Questão 02

- SGDClassifier

TF-IDF

		Previu	
		0	1
Verdadeiro	0	514	1
	1	2	151

	precision	recall	f1-score	support
0	1.00	1.00	1.00	515
1	0.99	0.99	0.99	153
accuracy			1.00	668
macro avg	0.99	0.99	0.99	668
weighted avg	1.00	1.00	1.00	668

CountVectorizer

		Previu	
		0	1
Verdadeiro	0	514	1
	1	2	151

	precision	recall	f1-score	support
0	0.99	1.00	1.00	515
1	0.99	0.98	0.99	153
accuracy			0.99	668
macro avg	0.99	0.99	0.99	668
weighted avg	0.99	0.99	0.99	668

Questão 02

d) Compare

Classificador	Vetorizador	F1-Score
Logistic Regression	CountVectorizer	0.9978
	TF-IDF	0.9508
Naive-Bayes	CountVectorizer	0.9652
	TF-IDF	0.9647
SGD	CountVectorizer	0.9914
	TF-IDF	0.9936

Questão 03

Realize um comparativo entre os métodos LDA, SVD e NMF para realizar modelagem de tópicos. Para cada método:

- a) empiricamente, tente identificar uma quantidade de tópicos adequada para seu problema. Exiba resultados que justifiquem a quantidade de tópicos escolhida;
- b) identifique as 5 palavras mais relevantes de cada tópico;
- c) identifique os 3 tópicos mais relevantes de 5 documentos quaisquer (você pode representar os tópicos por suas 5 palavras mais relevantes);
- d) qual método apresentou melhores resultados, na sua opinião? Justifique com resultados/dados.

Questão 03

- 5 tópicos é o número que trouxe os melhores resultados devido ao dataset original conter 5 classes.

NMF	Tópico 0	play, gam, win, england, champ
	Tópico 1	mr, labo elect, blair, party
	Tópico 2	us, mobl, phon, peopl, technolog
	Tópico 3	film, award, best, nomin, osc
	Tópico 4	econom, grow, bank, said, rat

LDA	Tópico 0	fairmont, savoy, alwalee, print alwalee, simpson strand
	Tópico 1	boers, deutsch boers, euronext, lse, fiat
	Tópico 2	said, us, mr, play, gam
	Tópico 3	pric, halifax, niss, lend, hous pric
	Tópico 4	lvmh, fal group, lacroix, christian lacroix, haut

SVD	Tópico 0	said, mr, us, year, would
	Tópico 1	mr, labo, elect, blair, party
	Tópico 2	mobl, us, phon, econom, firm
	Tópico 3	film, award, best, nomin, osc
	Tópico 4	econom, bank, rat, grow, pric

Questão 03

- Título: “Google launches TV search service”;
- Artigo: “The net search giant Google has launched a search service that lets people look for TV programmes...”;
- Rótulo: tech

NMF	Tópico 0	play, gam, win, england, champ	0.
	Tópico 1	mr, labo elect, blair, party	0.
	Tópico 2	us, mobl, phon, peopl, technolog	0.9314
	Tópico 3	film, award, best, nomin, osc	0.0685
	Tópico 4	econom, grow, bank, said, rat	0.

LDA	Tópico 0	fairmont, savoy, alwalee, print alwalee, simpson strand	0.0125
	Tópico 1	boers, deutsch boers, euronext, lse, fiat	0.0125
	Tópico 2	said,, us,, mr, play, gam	0.9496
	Tópico 3	pric, halifax, niss, lend, hous pric	0.0125
	Tópico 4	lvmh, fal group, lacroix, christian lacroix, haut	0.0125

SVD	Tópico 0	said, mr, us, year, would	0.3920
	Tópico 1	mr, labo, elect, blair, party	0.0668
	Tópico 2	mobl, us, phon, econom, firm	0.3128
	Tópico 3	film, award, best, nomin, osc	0.2282
	Tópico 4	econom, bank, rat, grow, pric	0.0000

Questão 03

- Título: “Highbury tunnel players in clear”;
- Artigo: “The Football Association has said it will not be bringing charges over the tunnel incident prior to...”;
- Rótulo: sport

NMF	Tópico 0	play, gam, win, england, champ	0.8744
	Tópico 1	mr, labo elect, blair, party	0.0396
	Tópico 2	us, mobl, phon, peopl, technolog	0.0859
	Tópico 3	film, award, best, nomin, osc	0.0000
	Tópico 4	econom, grow, bank, said, rat	0.0000

LDA	Tópico 0	fairmont, savoy, alwalee, print alwalee, simpson strand	0.0132
	Tópico 1	boers, deutsch boers, euronext, lse, fiat	0.0132
	Tópico 2	said,, us,, mr, play, gam	0.9471
	Tópico 3	pric, halifax, niss, lend, hous pric	0.0132
	Tópico 4	lvmh, fal group, lacroix, christian lacroix, haut	0.0132

SVD	Tópico 0	said, mr, us, year, would	0.7084
	Tópico 1	mr, labo, elect, blair, party	0.0800
	Tópico 2	mobl, us, phon, econom, firm	0.0418
	Tópico 3	film, award, best, nomin, osc	0.0000
	Tópico 4	econom, bank, rat, grow, pric	0.1696

Questão 03

- Título: “Carry On star Patsy Rowlands dies”;
- Artigo: “Actress Patsy Rowlands, known to millions for her roles in the Carry On films, has died at the age...”;
- Rótulo: entertainment

NMF	Tópico 0	play, gam, win, england, champ	0.1326
	Tópico 1	mr, labo elect, blair, party	0.0404
	Tópico 2	us, mobl, phon, peopl, technolog	0.0676
	Tópico 3	film, award, best, nomin, osc	0.7592
	Tópico 4	econom, grow, bank, said, rat	0.

LDA	Tópico 0	fairmont, savoy, alwalee, print alwalee, simpson strand	0.0123
	Tópico 1	boers, deutsch boers, euronext, lse, fiat	0.0123
	Tópico 2	said,, us,, mr, play, gam	0.9506
	Tópico 3	pric, halifax, niss, lend, hous pric	0.0123
	Tópico 4	lvmh, fal group, lacroix, christian lacroix, haut	0.0123

SVD	Tópico 0	said, mr, us, year, would	0.4272
	Tópico 1	mr, labo, elect, blair, party	0.
	Tópico 2	mobl, us, phon, econom, firm	0.0085
	Tópico 3	film, award, best, nomin, osc	0.3722
	Tópico 4	econom, bank, rat, grow, pric	0.1919

Questão 03

- Título: “Lesotho textile workers lose jobs”;
- Artigo: “Six foreign-owned textile factories have closed in Lesotho, leaving 6,650 garment workers jobless...”;
- Rótulo: business

NMF	Tópico 0	play, gam, win, england, champ	0.0759
	Tópico 1	mr, labo elect, blair, party	0.1480
	Tópico 2	us, mobl, phon, peopl, technolog	0.1501
	Tópico 3	film, award, best, nomin, osc	0.
	Tópico 4	econom, grow, bank, said, rat	0.6258

LDA	Tópico 0	fairmont, savoy, alwalee, print alwalee, simpson strand	0.0120
	Tópico 1	boers, deutsch boers, euronext, lse, fiat	0.0120
	Tópico 2	said,, us,, mr, play, gam	0.9516
	Tópico 3	pric, halifax, niss, lend, hous pric	0.0120
	Tópico 4	lvmh, fal group, lacroix, christian lacroix, haut	0.0120

SVD	Tópico 0	said, mr, us, year, would	0.4827
	Tópico 1	mr, labo, elect, blair, party	0.0811
	Tópico 2	mobl, us, phon, econom, firm	0.2063
	Tópico 3	film, award, best, nomin, osc	0.
	Tópico 4	econom, bank, rat, grow, pric	0.2297

Questão 03

- Título: “Ferguson hails Man Utd's resolve”;
- Artigo: “Manchester United's Alex Ferguson has praised his players' gutsy performance in the 1-0 win at Aston...”;
- Rótulo: sport

NMF	Tópico 0	play, gam, win, england, champ	0.9981
	Tópico 1	mr, labo elect, blair, party	0.
	Tópico 2	us, mobl, phon, peopl, technolog	0.
	Tópico 3	film, award, best, nomin, osc	0.0018
	Tópico 4	econom, grow, bank, said, rat	0.

LDA	Tópico 0	fairmont, savoy, alwalee, print alwalee, simpson strand	0.0162
	Tópico 1	boers, deutsch boers, euronext, lse, fiat	0.0162
	Tópico 2	said,, us,, mr, play, gam	0.9349
	Tópico 3	pric, halifax, niss, lend, hous pric	0.0162
	Tópico 4	lvmh, fal group, lacroix, christian lacroix, haut	0.0162

SVD	Tópico 0	said, mr, us, year, would	0.6847
	Tópico 1	mr, labo, elect, blair, party	0.0399
	Tópico 2	mobl, us, phon, econom, firm	0.
	Tópico 3	film, award, best, nomin, osc	0.0497
	Tópico 4	econom, bank, rat, grow, pric	0.2255

Questão 03

- d) qual método apresentou melhores resultados, na sua opinião?

Como os resultados mostram, podemos ver que o NMF trouxe uma melhor separação em tópicos do que os outros dois algoritmos. O LDA teve o pior desempenho, não conseguindo discernir muito bem as características de cada tópico. Cada tópico escolhido pelo NMF pode ser associado a uma classe real do dataset. Por exemplo, a notícia número 1 tem as palavras mais associadas ao tópico de tecnologia (tópico 2) e de fato é uma notícia que é classificada como notícia de tecnologia no dataset original.

Questão 04

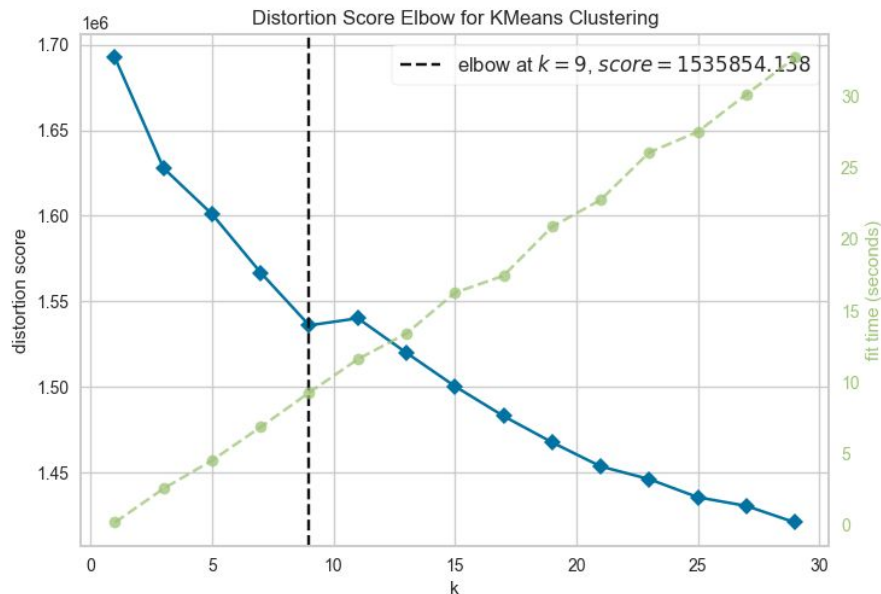
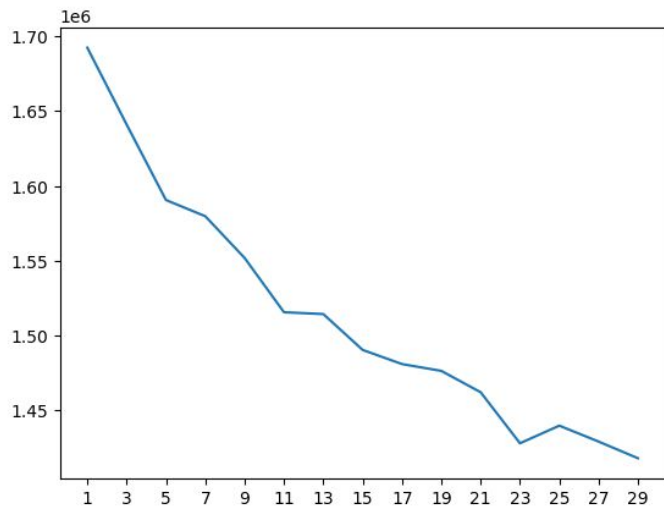
Realize um agrupamento dos dados usando a representação do CountVectorizer, seguindo os seguintes passos:

a) Aplique o algoritmo PCA, preservando 95% da variância nos dados. Qual a dimensão resultante dos dados projetados?

Originalmente, com nossa base de dados, o CountVectorizer gera um vetor de 305530 de tamanho. Com o PCA preservando 95% da variância dos dados, passamos a ter um vetor de 1529 de tamanho para representar nossos dados.

Questão 04

b) Aplique um algoritmo k-means nos dados projetados, tentando usar o método elbow para encontrar o valor de k ótimo.



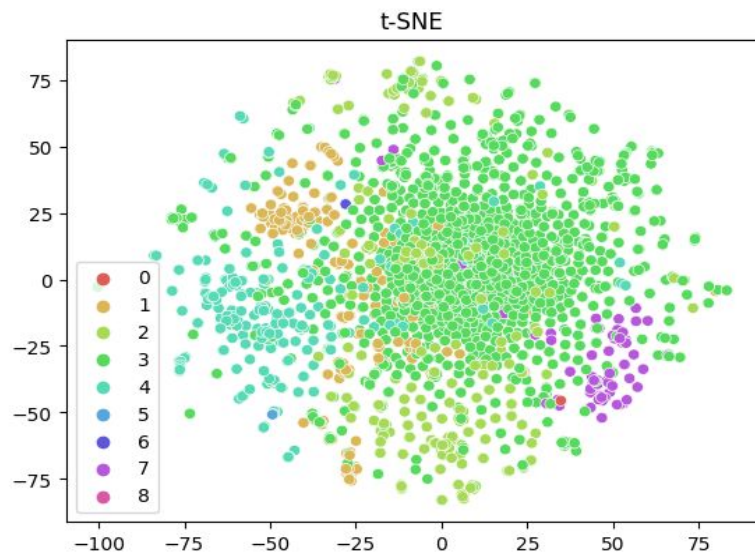
Questão 05

Nesta questão você deve aplicar métodos de projeção multidimensional para visualizar os dados da segunda questão no espaço visual.

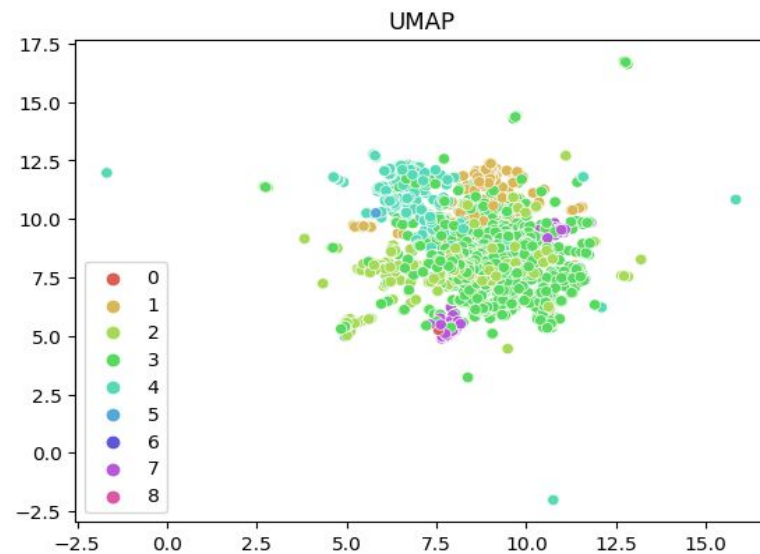
a) aplique os métodos de projeção multidimensional t-SNE e UMAP na representação CountVectorizer e plote os gráficos das projeções resultantes, colorindo os pontos de acordo com os grupos obtidos pelo k-means (questão 4b).

Questão 05

*TSNE(n_components=2, learning_rate='auto', init='random',
perplexity=3, square_distances=True)*



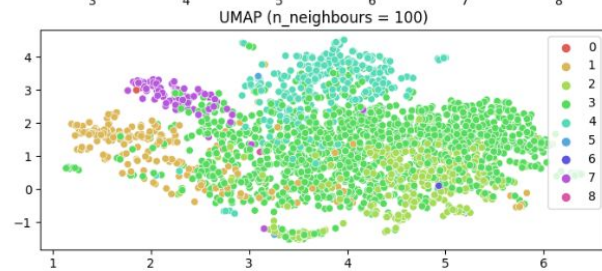
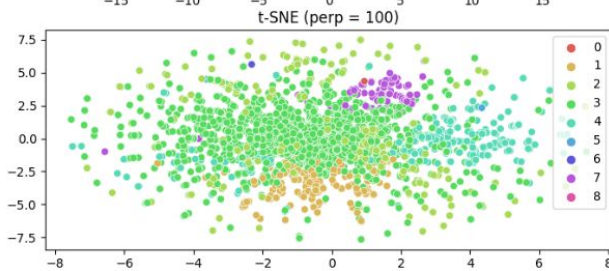
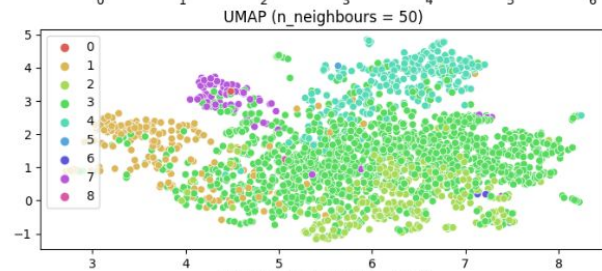
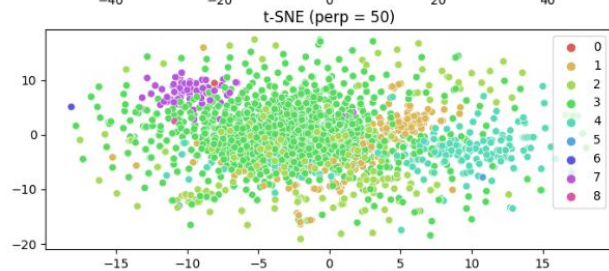
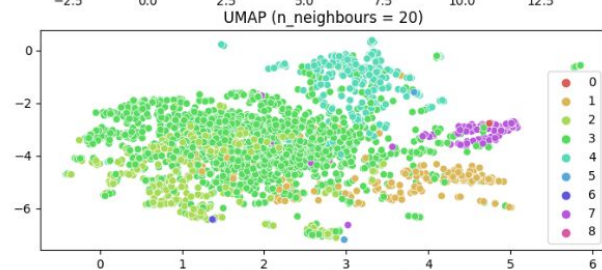
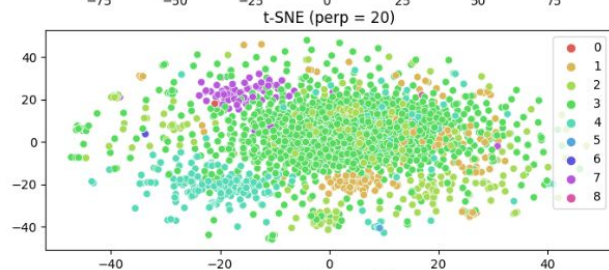
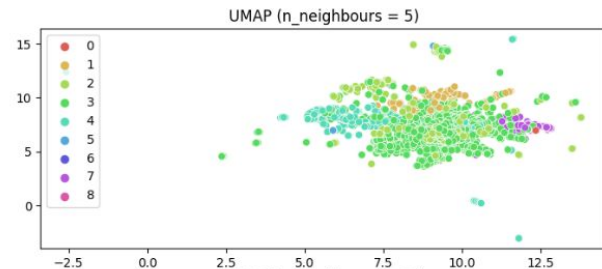
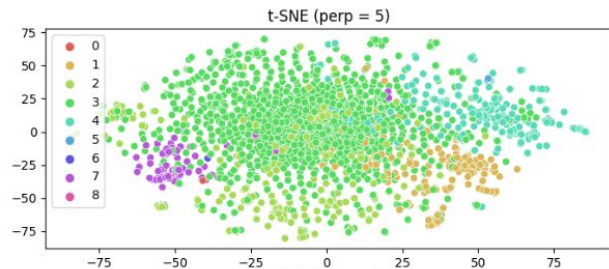
UMAP(n_neighbors=5)



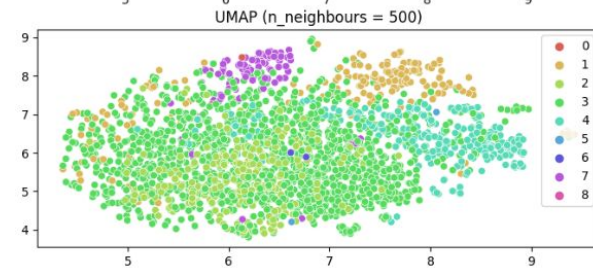
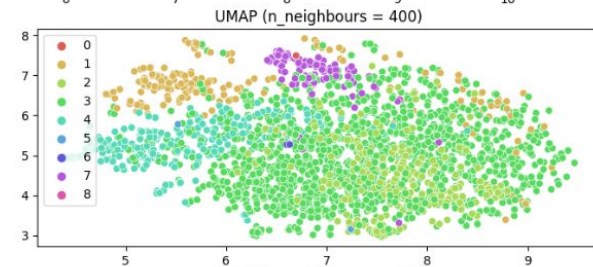
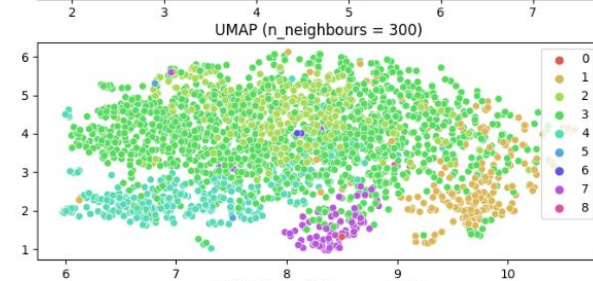
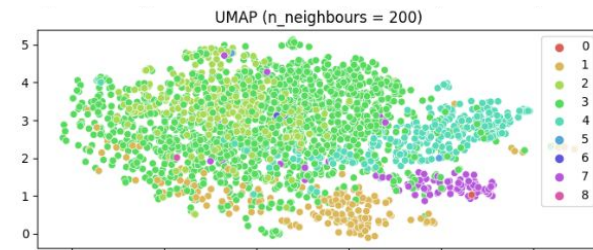
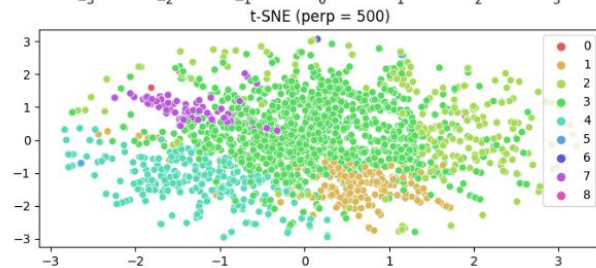
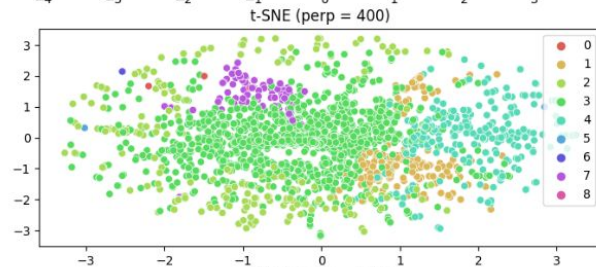
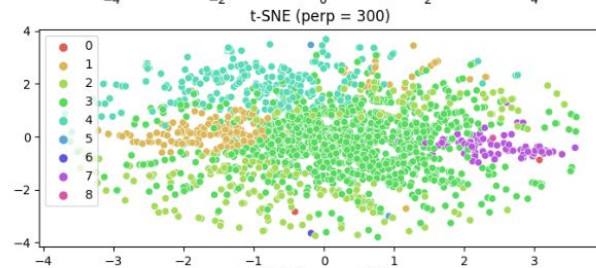
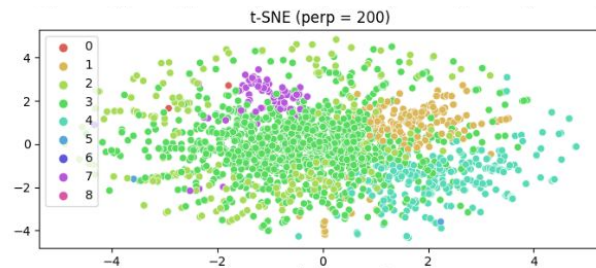
Questão 05

b) Experimente variar os hiperparâmetros perplexity do t-SNE e n neighbors do UMAP. O que acontece com as projeções quando estes parâmetros são calibrados para valores menores ou maiores do que seus valores padrão?

Questão 05



Questão 05



Questão 05

b) Experimente variar os hiperparâmetros perplexity do t-SNE e n neighbors do UMAP. O que acontece com as projeções quando estes parâmetros são calibrados para valores menores ou maiores do que seus valores padrão?

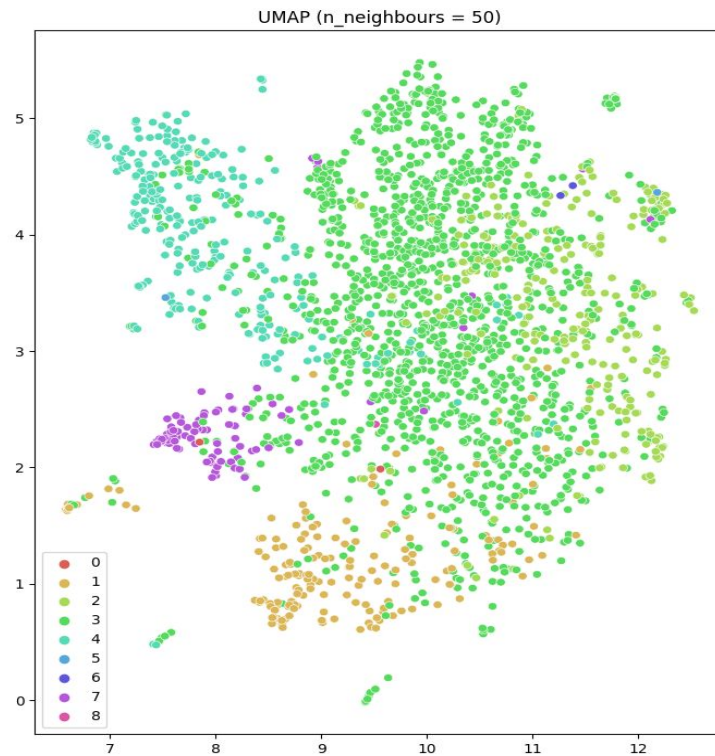
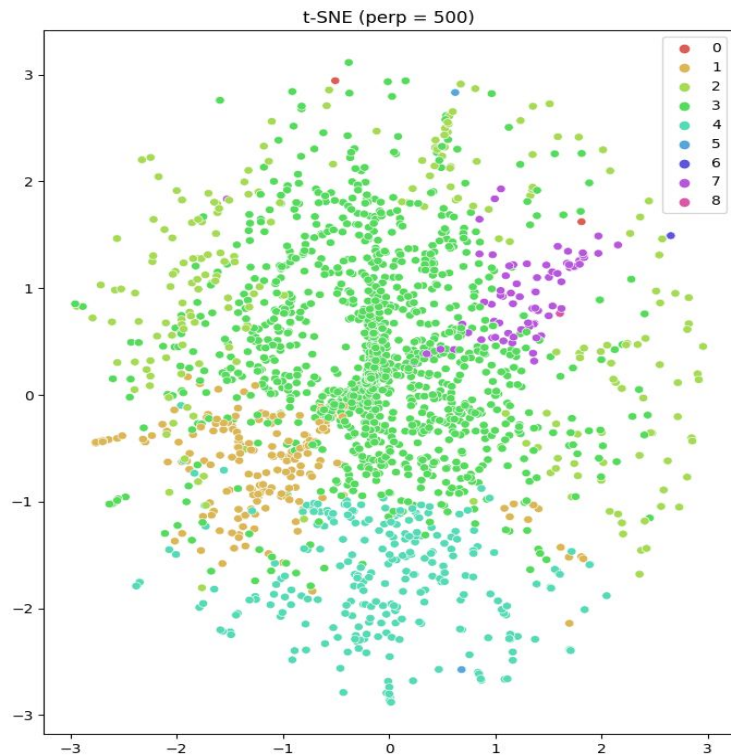
Podemos notar que ao aumentar o número de vizinhos no UMAP, ele começa a perder algumas informações mais detalhadas. Por exemplo, quando `n_neighbours = 500` as classes 1, 2 e 3 acabam se misturando bastante. Enquanto que com `n_neighbours=50` podemos visualizar melhor uma separação nas bordas dos clusters principais. Porém, quando diminuimos muito esse valor de `n_neighbours`, os pontos projetados ficam muito próximos no espaço e os grupos acabam tendo poucas diferenças.

Já quanto ao valor de perplexity do t-SNE, quanto menor seu valor, mais espalhado ficam os dados projetados. A medida que aumentamos, começamos a ver dados mais globais. Podemos notar que para `perplexity=20`, a classe 3 fica mais espalhada no espaço. Porém ao plotarmos para `perplexity=500`, conseguimos ver que a classe 3 é menos espalhada e consegue ser melhor distinguida.

Questão 05

c) Usando os melhores valores encontrados para perplexity e n neighbors, compare o tempo de execução e a qualidade visual da projeção dos 2 métodos de projeção multidimensional (t-SNE e UMAP). Qual abordagem se saiu melhor em cada um desses aspectos?

Questão 05



Questão 05

c) Usando os melhores valores encontrados para perplexity e n neighbors, compare o tempo de execução e a qualidade visual da projeção dos 2 métodos de projeção multidimensional (t-SNE e UMAP). Qual abordagem se saiu melhor em cada um desses aspectos?

O t-SNE levou 49.35 segundos, enquanto o UMAP levou 5.40 segundos. Logo, o tempo para rodar o t-SNE é muito maior que o UMAP. Analisando os plots, conseguimos notar que o UMAP consegue ter um melhor agrupamento dos dados quando comparado com o t-SNE. As classes 2 e 3 estão bastante esparsas no espaço, enquanto que no UMAP as classes conseguem ser melhor distinguidas.

Lista 3

Questão 02

Resolva novamente a segunda questão da 2a lista e compare com os resultados obtidos anteriormente:

a) Aplicando a representação vetorial Doc2Vec combinado com os classificadores usados anteriormente.

Questão 02

- Doc2Vec

Logistic Regression

		Previu	
		0	1
Verdadeiro	0	528	0
	1	2	138

	precision	recall	f1-score	support
0	1.00	1.00	1.00	528
1	1.00	0.99	0.99	140
accuracy			1.00	668
macro avg	1.00	0.99	1.00	668
weighted avg	1.00	1.00	1.00	668

Naive-Bayes

		Previu	
		0	1
Verdadeiro	0	386	142
	1	19	121

	precision	recall	f1-score	support
0	0.95	0.73	0.83	528
1	0.46	0.86	0.60	140
accuracy			0.76	668
macro avg	0.71	0.80	0.71	668
weighted avg	0.85	0.76	0.78	668

SGD

		Previu	
		0	1
Verdadeiro	0	528	0
	1	5	135

	precision	recall	f1-score	support
0	0.99	1.00	1.00	528
1	1.00	0.96	0.98	140
accuracy			0.99	668
macro avg	1.00	0.98	0.99	668
weighted avg	0.99	0.99	0.99	668

Questão 02

b) Usando pelo menos duas arquiteturas de redes neurais que utilizem camadas Embedding, convolucionais e LSTM.

Rede	Acurácia	F1-Score
Modelo 1 (Conv1D)	98.05%	98.21%
Modelo 2 (LSTM)	99.55%	99.48%

Questão 02

Comparando

Classificador	Vetorizador	F1-Score
Logistic Regression	CountVectorizer	0.9978
	TF-IDF	0.9508
	Doc2Vec	0.9954
Naive-Bayes	CountVectorizer	0.9652
	TF-IDF	0.9647
	Doc2Vec	0.7139
SGD	CountVectorizer	0.9914
	TF-IDF	0.9936
	Doc2Vec	0.9885
Modelo 1 (Conv1D)		0.9821
Modelo 2 (LSTM)		0.9948

Questão 04

4. Usando sua base de textos e a biblioteca spaCy, realize as seguintes tarefas:

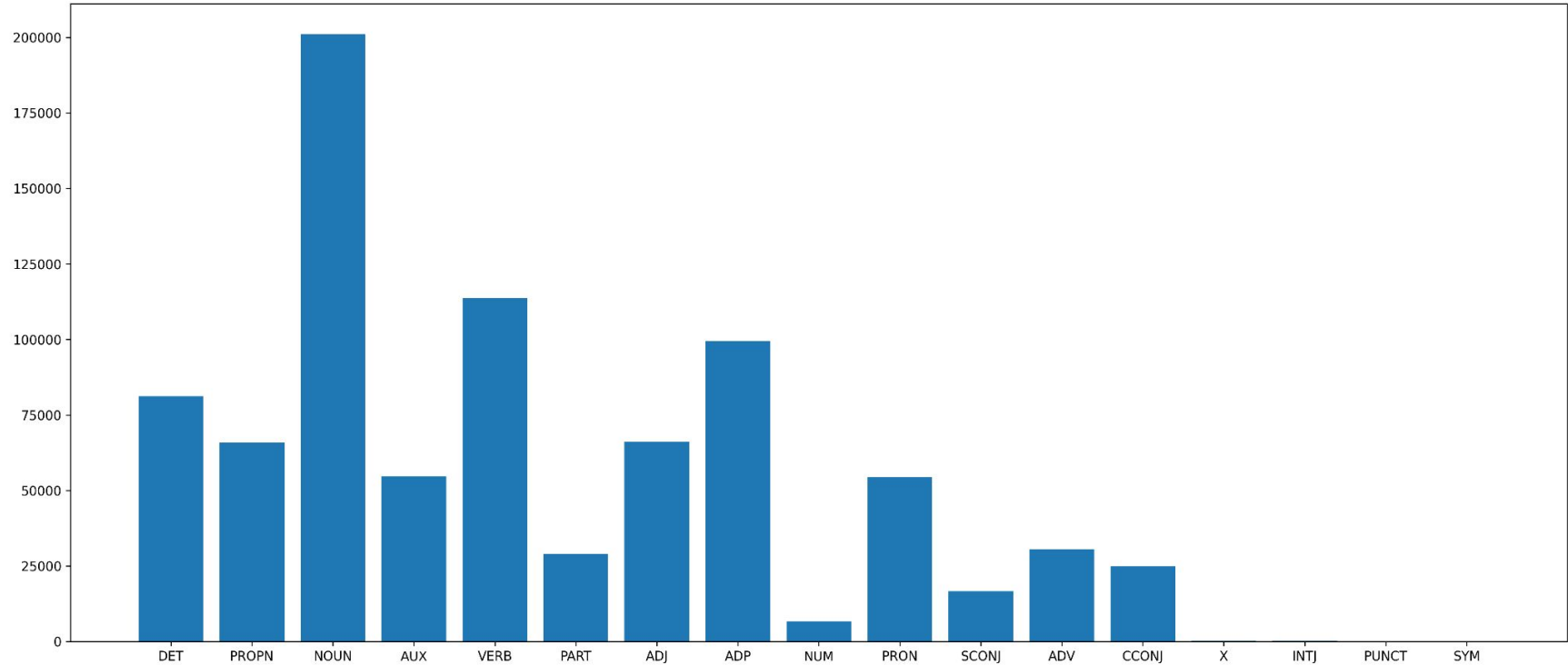
a) Extraia as etiquetas gramaticais (POS) de cada token do seu textos.

```
print(counters)
```

```
{'DET': 81235, 'PROPN': 65930, 'NOUN': 201140, 'AUX': 54596, 'VERB': 113707, 'PART': 29001, 'ADJ': 66104, 'ADP': 99444, 'NUM': 6645, 'PRON': 54440, 'SCONJ': 16650, 'ADV': 30598, 'CCONJ': 25002, 'X': 287, 'INTJ': 295, 'PUNCT': 7, 'SYM': 8}
```

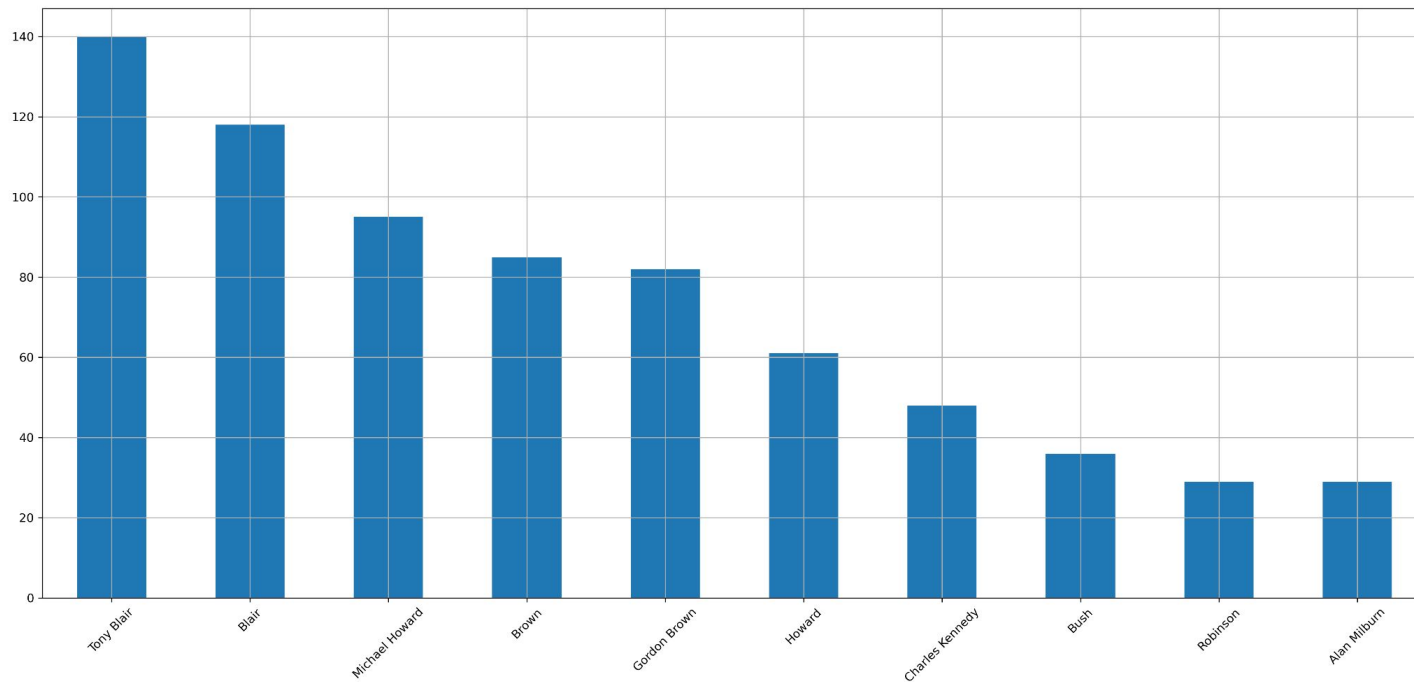
Questão 04

b) Calcule e plote um gráfico com as frequências de cada tipo gramatical.



Questão 04

d) Identifique e liste as pessoas mais frequentes nos seus textos. Você só deve contar cada entidade 1 vez por documento.



Lista 4

Questão 01

1. Estude o notebook “A Visual Notebook to Using BERT for the First Time.ipynb”, anexo a essa lista.

a) Resolva o mesmo problema de classificação da segunda questão da segunda lista, usando uma combinação de DistilBERT com os três classificadores usados na segunda lista.

Questão 01

- BERT

Logistic Regression

		Previu	
		0	1
Verdadeiro	0	340	3
	1	0	102

	precision	recall	f1-score	support
0	1.00	0.99	1.00	343
1	0.97	1.00	0.99	102
accuracy			0.99	445
macro avg	0.99	1.00	0.99	445
weighted avg	0.99	0.99	0.99	445

Naive-Bayes

		Previu	
		0	1
Verdadeiro	0	339	4
	1	6	96

	precision	recall	f1-score	support
0	0.98	0.99	0.99	343
1	0.96	0.94	0.95	102
accuracy			0.98	445
macro avg	0.97	0.96	0.97	445
weighted avg	0.98	0.98	0.98	445

SGD

		Previu	
		0	1
Verdadeiro	0	330	13
	1	0	102

	precision	recall	f1-score	support
0	1.00	0.96	0.98	343
1	0.89	1.00	0.94	102
accuracy			0.97	445
macro avg	0.94	0.98	0.96	445
weighted avg	0.97	0.97	0.97	445

Questão 01

b) Compare todos os resultados.

Classificador	Vetorizador	F1-Score
Logistic Regression	CountVectorizer	0.9978
	TF-IDF	0.9508
	Doc2Vec	0.9954
	BERT	0.9905
Naive-Bayes	CountVectorizer	0.9652
	TF-IDF	0.9647
	Doc2Vec	0.7139
	BERT	0.9679
SGD	CountVectorizer	0.9914
	TF-IDF	0.9936
	Doc2Vec	0.9885
	BERT	0.9603
Modelo 1 (Conv1D)		0.9821
Modelo 2 (LSTM)		0.9948

Questão 02

2. Estude e pesquise sobre o BERTopic, uma adaptação do BERT para modelagem de tópicos

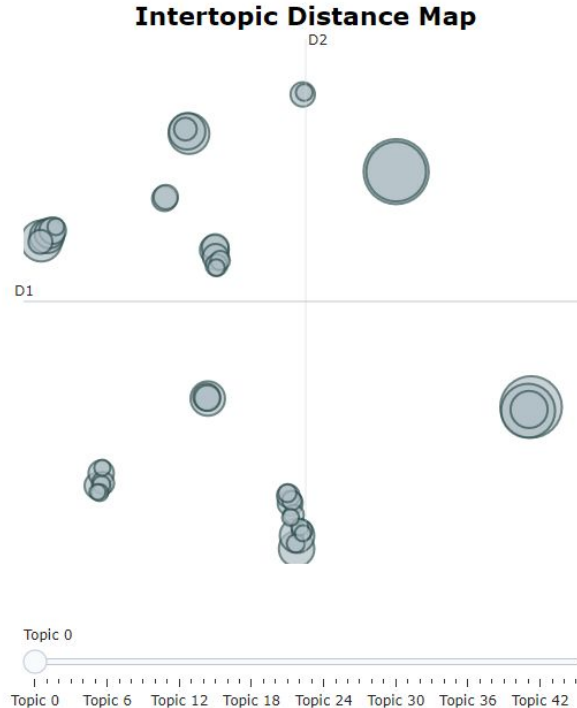
a) Extraia os tópicos de sua base, exibindo as informações dos tópicos (palavras mais relevantes)

Questão 02

Tópico 0	('england', '0.019'), ('club', '0.017'), ('game', '0.016'), ('side', '0.014'), ('wales', '0.013'), ('players', '0.013')
Tópico 1	('film', '0.054'), ('best', '0.030'), ('actor', '0.021'), ('films', '0.020'), ('director', '0.019'), ('awards', '0.018')
Tópico 2	('music', '0.031'), ('band', '0.030'), ('album', '0.029'), ('song', '0.023'), ('best', '0.019'), ('rock', '0.017')
Tópico 3	('race', '0.035'), ('olympic', '0.035'), ('indoor', '0.030'), ('champion', '0.027'), ('championships', '0.026'), ('holmes', '0.024')
Tópico 4	('police', '0.026'), ('rights', '0.020'), ('government', '0.019'), ('human', '0.019'), ('said', '0.017'), ('law', '0.017')
Tópico 5	('ebbers', '0.032'), ('fraud', '0.031'), ('company', '0.026'), ('worldcom', '0.025'), ('mr', '0.025'), ('sec', '0.021')
Tópico 6	('technology', '0.023'), ('gadgets', '0.021'), ('digital', '0.020'), ('technologies', '0.018'), ('people', '0.017'), ('devices', '0.015')
Tópico 7	('roddick', '0.058'), ('seed', '0.037'), ('nadal', '0.036'), ('open', '0.034'), ('set', '0.030'), ('federer', '0.028')
Tópico 8	('blair', '0.049'), ('labour', '0.047'), ('mr', '0.046'), ('brown', '0.045'), ('election', '0.036'), ('prime', '0.034')
Tópico 9	('spam', '0.046'), ('mail', '0.044'), ('virus', '0.038'), ('site', '0.030'), ('attacks', '0.029'), ('security', '0.028')
Tópico 10	('mobile', '0.059'), ('phone', '0.042'), ('phones', '0.034'), ('technology', '0.022'), ('people', '0.022'), ('mobiles', '0.021')

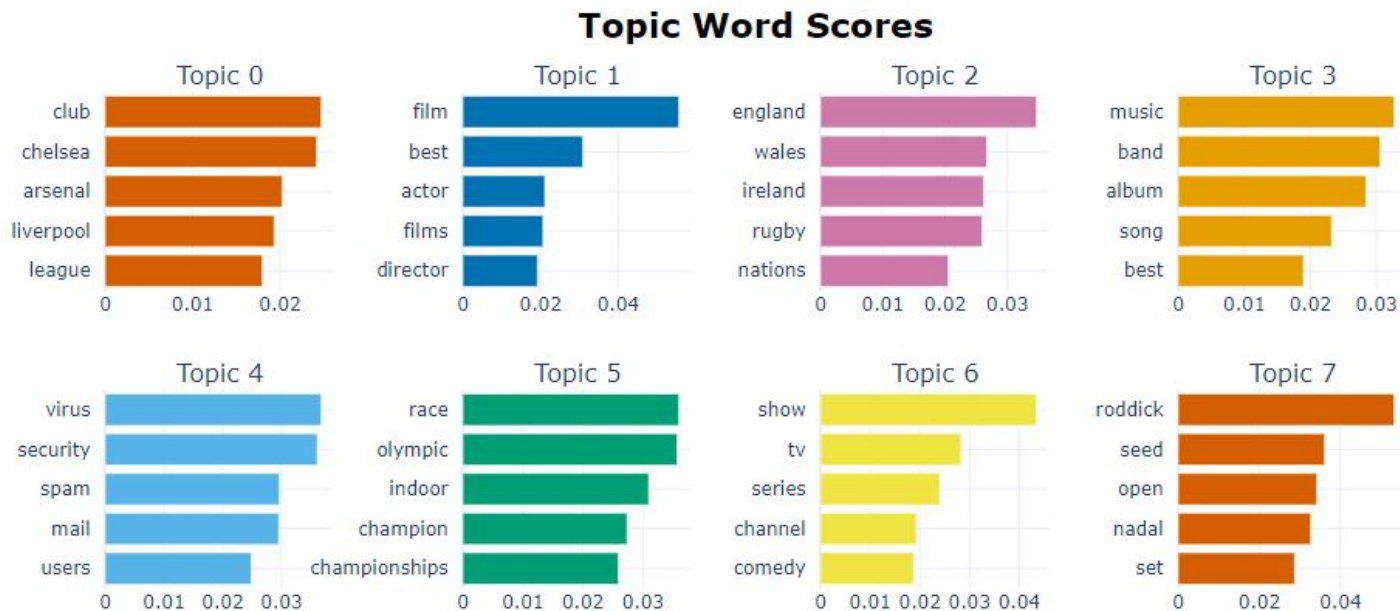
Questão 02

b) Exiba visualizações com gráficos de barra e usando visualize topics().



Questão 02

b) Exiba visualizações com gráficos de barra e usando visualize topics().



Obrigado!