

Relatório TP4

Ellen Christina Amaral Santana - 12011BSI208

Exercício 1

Com o tf-idf os resultados ficaram bem semelhantes, apenas o último elemento do array teve um predict diferente e a acurácia diminuiu 0.01 em relação ao inicial.

```
modelo.predict(freq_testes)
```

```
array(['Negativo', 'Neutro', 'Negativo', 'Neutro', 'Positivo'],  
      dtype='<U8')
```

```
metrics.accuracy_score(classes, resultados)
```

```
0.8768142456397121
```

O desempenho por classe teve pequenas alterações: precisão pouco maior no negativo e no neutro, menor no positivo; recall menor em neutro, maior em positivo.

```
print(metrics.classification_report(classes, resultados))
```

	precision	recall	f1-score	support
Negativo	0.88	0.93	0.90	2446
Neutro	0.83	0.79	0.81	2453
Positivo	0.91	0.90	0.91	3300
accuracy			0.88	8199
macro avg	0.87	0.87	0.87	8199
weighted avg	0.88	0.88	0.88	8199

Quantidade de acerto por classe, igual à da análise inicial.

```
print (pd.crosstab(classes, resultados, rownames=['Real'], colnames=['Predito'], margins=True))
```

Predito	Negativo	Neutro	Positivo	All
Real				
Negativo	2281	143	22	2446
Neutro	257	1933	263	2453
Positivo	61	264	2975	3300
All	2599	2340	3260	8199

Exercício 2

Usando TF-IDF com bigrama, temos uma acurácia maior que a análise inicial.

```
metrics.accuracy_score(classes,resultados)
```

```
0.8896206854494451
```

A classificação por classe em relação à análise inicial ficou: precisão menor para negativo, maior para neutro, menor para positivo; recall pouco maior para negativo, menor para neutro e maior para positivo.

```
print(metrics.classification_report(classes,resultados))
```

	precision	recall	f1-score	support
Negativo	0.87	0.94	0.90	2446
Neutro	0.85	0.82	0.83	2453
Positivo	0.94	0.90	0.92	3300
accuracy			0.89	8199
macro avg	0.88	0.89	0.89	8199
weighted avg	0.89	0.89	0.89	8199

E a quantidade de acerto por classe aumentou para negativo, neutro e positivo.

```
print (pd.crosstab(classes, resultados, rownames=['Real'], colnames=['Predito'], margins=True))
```

Predito	Negativo	Neutro	Positivo	All
Real				
Negativo	2300	130	16	2446
Neutro	268	2009	176	2453
Positivo	78	237	2985	3300
All	2646	2376	3177	8199

Exercício 4

Removendo as stopwords do modelo tf-idf com bigrams, temos:

```
modelo.predict(freq_testes)
```

```
array(['Neutro', 'Neutro', 'Negativo', 'Neutro', 'Neutro'], dtype='<U8')
```

```
metrics.accuracy_score(classes,resultados)
```

```
0.871447737528967
```

```
print(metrics.classification_report(classes,resultados))
```

	precision	recall	f1-score	support
Negativo	0.87	0.91	0.89	2446
Neutro	0.84	0.78	0.81	2453
Positivo	0.90	0.91	0.91	3300
accuracy			0.87	8199
macro avg	0.87	0.87	0.87	8199
weighted avg	0.87	0.87	0.87	8199

```
print (pd.crosstab(classes, resultados, rownames=['Real'], colnames=['Predito'], margins=True))
```

Predito	Negativo	Neutro	Positivo	All
Real				
Negativo	2219	163	64	2446
Neutro	258	1910	285	2453
Positivo	85	199	3016	3300
All	2562	2272	3365	8199

Resultados um pouco piores que o modelo com as stopwords.