

# Relatório

Otávio Malta Borges - 12011BSI291

## Ex1

Com o TF-IDF e 1-gram, a acurácia diminuiu 0,01 do TF com 1-gram e 0,02 do TF com 2-gram, cujo obteve o melhor resultado. É possível notar este resultado ao se analisar a precisão de cada classe (Positivo, Neutro ou Negativo). O TF-IDF com 1-gram, por exemplo, obteve a menor precisão em todas as classes, enquanto o TF com 2-gram obteve as maiores.

Porém, ao se analisar a quantidade de acertos por classe, houve um aumento em nas classes Negativo e Positivo.

	['Negativo'	'Neutro'	'Negativo'	'Neutro'	'Positivo']
	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	
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	
Acurácia: 0.8768142456397121					

Naive Bayes + TF-IDF + 1-gram

	precision	recall	f1-score	support	
Negativo	0.89	0.93	0.91	2446	
Neutro	0.80	0.84	0.82	2453	
Positivo	0.95	0.88	0.91	3300	
accuracy			0.88	8199	
macro avg	0.88	0.88	0.88	8199	
weighted avg	0.89	0.88	0.88	8199	
Predito	Negativo	Neutro	Positivo	All	
Real					
Negativo	2275	162	9	2446	
Neutro	240	2067	146	2453	
Positivo	45	356	2899	3300	
All	2560	2585	3054	8199	
Acurácia: 0.8831564824978656					

Naive Bayes + TF + 1-gram

	precision	recall	f1-score	support	
Negativo	0.91	0.93	0.92	2446	
Neutro	0.80	0.89	0.84	2453	
Positivo	0.97	0.88	0.92	3300	
accuracy			0.90	8199	
macro avg	0.89	0.90	0.89	8199	
weighted avg	0.90	0.90	0.90	8199	
Predito	Negativo	Neutro	Positivo	All	
Real					
Negativo	2265	179	2	2446	
Neutro	181	2177	95	2453	
Positivo	43	357	2900	3300	
All	2489	2713	2997	8199	
Acurácia: 0.8954750579338944					

Naive Bayes + TF + 2-gram

## Ex2

O TF-IDF com 2-gram obteve uma acurácia maior que com 1-gram, com 0,01 a mais (0.889). Porém ainda menor que o cenário com Naive Bayes, TF e 2-gram.

Obteve precisão maior para as classes Neutro e Positivo e menor para Negativo.

A quantidade de acertos por classe aumentou para as classes Negativo e Neutro

	['Negativo' 'Neutro' 'Negativo' 'Neutro' 'Neutro']				
	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	
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	
Acurácia: 0.8896206854494451					

Naive Bayes + TF-IDF + 2-gram

## Ex3

Até agora, o melhor resultado encontrado com relação a BoW foi o TF + 2-gram. Para este exercício, foram utilizados os seguintes algoritmos de classificação: Árvore de Decisão, Floresta Aleatória e Máquinas de Vetores de Suporte.

O primeiro obteve a menor acurácia dos 3 (0.872).

O desempenho de cada uma das classes foram os menores entre os algoritmos (exceto para a classe Neutro), obtendo uma média de 0.87.

Entretanto, a quantidade de acertos por classe foi a maior para Negativo e Positivo.

Árvore de Decisão					
	['Negativo' 'Negativo' 'Neutro' 'Neutro' 'Negativo']				
	precision	recall	f1-score	support	
Negativo	0.89	0.90	0.89	2446	
Neutro	0.80	0.82	0.81	2453	
Positivo	0.92	0.89	0.90	3300	
accuracy			0.87	8199	
macro avg	0.87	0.87	0.87	8199	
weighted avg	0.87	0.87	0.87	8199	
Predito	Negativo	Neutro	Positivo	All	
Real					
Negativo	2199	181	66	2446	
Neutro	232	2022	199	2453	
Positivo	38	332	2930	3300	
All	2469	2535	3195	8199	
Acurácia: 0.8721795340895231					

Arvore de Decisão + TF + 2-gram,

Floresta Aleatória					
	'Neutro'	'Neutro'	'Neutro'	'Neutro'	'Negativo'
	precision		recall	f1-score	support
Negativo	0.97	0.88	0.92	2446	
Neutro	0.75	0.94	0.83	2453	
Positivo	0.97	0.85	0.91	3300	
accuracy			0.89	8199	
macro avg	0.90	0.89	0.89	8199	
weighted avg	0.90	0.89	0.89	8199	
Predito	Negativo	Neutro	Positivo	All	
Real					
Negativo	2143	300	3	2446	
Neutro	61	2318	74	2453	
Positivo	12	487	2801	3300	
All	2216	3105	2878	8199	
Acurácia: 0.8857177704598121					

Floresta Aleatória + TF + 2-gram

Floresta Aleatória obteve uma acurácia maior que o antecessor(0.885).

O desempenho obtido em cada uma das classes obteve uma média de 0.90. Porém, a precisão para a classe Neutro foi a menor entre os 3 algoritmos (0.75).

Sua quantidade de acertos por classe foi maior para Neutros e menor para as outras duas em relação à Árvore de Decisão.

Já a Máquinas de Vetores de Suporte obteve a maior acurácia entre as 3 (0.894). Entretanto, ainda obteve um resultado menor que o algoritmo Naive Bayes, mesmo que apenas 0.001 de diferença.

O desempenho em cada uma das classes obteve uma média de 0.90, enquanto o algoritmo Naive Bayes teve uma média de 0.89.

Em relação à quantidade de acertos por classes, o algoritmo não teve nenhum destaque.

SVM					
	'Neutro'	'Negativo'	'Neutro'	'Neutro'	'Negativo'
	precision		recall	f1-score	support
Negativo	0.95	0.90	0.92	2446	
Neutro	0.78	0.92	0.84	2453	
Positivo	0.97	0.88	0.92	3300	
accuracy			0.89	8199	
macro avg	0.90	0.90	0.89	8199	
weighted avg	0.91	0.89	0.90	8199	
Predito	Negativo	Neutro	Positivo	All	
Real					
Negativo	2191	251	4	2446	
Neutro	111	2253	89	2453	
Positivo	11	395	2894	3300	
All	2313	2899	2987	8199	
Acurácia: 0.8949871935601903					

Máquinas de Vetores de Suporte + TF + 2-gram

Nota-se que, desde o exercício 1, as precisões para a classe Neutro em todos os algoritmos foram sempre menores que as outras duas classes. A classe Positivo sempre foi a maior ou igual a Negativo, ou seja, a classe Negativo em nenhum dos casos obteve a menor nem a maior precisão.

## Ex4

Ao remover as Stopwords e refazer o experimento usando Naive Bayes com TF e 2-gram, foi obtido uma acurácia de 0.898, um aumento de 0,003.

	['Negativo'	'Negativo'	'Negativo'	'Neutro'	'Negativo']
	precision	recall	f1-score	support	
Negativo	0.90	0.93	0.92	2446	
Neutro	0.82	0.89	0.85	2453	
Positivo	0.97	0.88	0.92	3300	
accuracy			0.90	8199	
macro avg	0.90	0.90	0.90	8199	
weighted avg	0.90	0.90	0.90	8199	
Predito	Negativo	Neutro	Positivo	All	
Real					
Negativo	2279	161	6	2446	
Neutro	182	2182	89	2453	
Positivo	59	333	2908	3300	
All	2520	2676	3003	8199	
Acurácia: 0.8987681424563971					

Naive Bayes + TF + 2-gram (sem Stopwords)

	precision	recall	f1-score	support
Negativo	0.91	0.93	0.92	2446
Neutro	0.80	0.89	0.84	2453
Positivo	0.97	0.88	0.92	3300
accuracy			0.90	8199
macro avg	0.89	0.90	0.89	8199
weighted avg	0.90	0.90	0.90	8199
Predito	Negativo	Neutro	Positivo	All
Real				
Negativo	2265	179	2	2446
Neutro	181	2177	95	2453
Positivo	43	357	2900	3300
All	2489	2713	2997	8199
Acurácia: 0.8954750579338944				

Naive Bayes + TF + 2-gram (com Stopwords)

O desempenho obtido em cada uma das classes obteve uma média de 0.90, sendo maior que o experimento com Stopwords (0.89)

Sua quantidade de acertos por classe foi maior para Positivos e Negativos.

## Ex5

Ao utilizar o arquivo "reforma\_previdencia\_rotulado.csv" com TF, 2-gram, Naive Bayes e sem stopwords, o valor da acurácia caiu significativamente (0.6088).

O desempenho em cada uma das classes obteve uma média de 0.61, ou seja, 0.29 a menos que a mesma configuração com o arquivo anterior.

A quantidade de acertos por classe também diminuiu, visto que o arquivo anterior possuía 7581 tweets a mais, o que levou a queda da precisão das classes e da acurácia.

	['Negativo'	['Negativo'	['Negativo'	['Positivo'	['Positivo']
	precision	recall	f1-score	support	
Negativo	0.61	0.56	0.58	712	
Neutro	0.64	0.58	0.61	780	
Positivo	0.59	0.69	0.63	740	
accuracy			0.61	2232	
macro avg	0.61	0.61	0.61	2232	
weighted avg	0.61	0.61	0.61	2232	
Predito	Negativo	Neutro	Positivo	All	
Real					
Negativo	401	142	169	712	
Neutro	141	449	190	780	
Positivo	119	112	509	740	
All	661	703	868	2232	
Acurácia:	0.6088709677419355				

Naive Bayes + TF + 2-gram (sem Stopwords)

	['Negativo'	['Negativo'	['Negativo'	['Positivo'	['Positivo']
	precision	recall	f1-score	support	
Negativo	0.60	0.57	0.59	712	
Neutro	0.66	0.55	0.60	780	
Positivo	0.57	0.71	0.63	740	
accuracy			0.61	2232	
macro avg	0.61	0.61	0.61	2232	
weighted avg	0.61	0.61	0.61	2232	
Predito	Negativo	Neutro	Positivo	All	
Real					
Negativo	405	124	183	712	
Neutro	145	426	209	780	
Positivo	122	95	523	740	
All	672	645	915	2232	
Acurácia:	0.6066308243727598				

Naive Bayes + TF + 2-gram  
(comStopwords)

Após realizar diversos testes com as combinações de algoritmos de classificação e BoW, é notório que o resultado segue proporcional aos experimentos realizados com o arquivo anterior.

Para todos os cenários, o desempenho foi menor, resultado da diminuição de dados usados para treinar os modelos.

Além disso, a precisão da classe Neutro deixou de ser sempre a menor em todos os cenários. Nos experimentos realizados, as precisões menores oscilaram entre Positivo e Negativo.