



Detecção de Tweets Spam

Cláudio Torres Junior
Eloiza Rossetto dos Santos



Motivação

- Redes sociais estão cada vez mais presente em nossos dia-a-dias.
- O Twitter é uma das redes sociais mais utilizadas. Possui 396 milhões de usuários sendo que 206 milhões que acessam a rede diariamente.
- Por ser uma rede pública e de fácil acesso, cyber criminosos aproveitam de hashtags e trending topics para disseminar mensagens de phishing e spam.



Objetivos

- Identificar tweets SPAM de não-SPAM de maneira automática e rápida.
- Analisar tweets SPAM em discussões do Twitter brasileiro.



Base de dados

- Para o trabalho utilizamos uma base pública de tweets políticos do Twitter americano de 2018;
- A base possui ao todo 22.342 amostras rotuladas. Sendo 7443 (49,96%) da classe SPAM e 7554 (49,96%) da classe não-SPAM

Base de dados

	Tweet	following	followers	actions	is_retweet	location	Type	Unnamed: 7
0	Good Morning Love @LeeBrown_V	0.0	0.0	0.0	0.0	Pennsylvania, USA	Quality	NaN
1	'@realDonaldTrump @USNavy RIP TO HEROES'	42096.0	61060.0	5001.0	0.0	South Padre Island, Texas	Spam	NaN
2	Haven't been following the news but I understa...	0.0	0.0	NaN	0.0	Will never be broke ever again	Quality	NaN
3	pic.twitter.com/dy9q4fLhZ What to do with pap...	0.0	0.0	0.0	0.0	Mundo	Quality	NaN
4	#DidYouKnow ► Mahatma Gandhi made a brief visi...	17800.0	35100.0	NaN	0.0	Nottingham, England	Quality	NaN
...
14894	#AllWentWrongWhen I told my hair stylist to "g...	695.0	533.0	868.0	1.0	United States	Spam	NaN
14895	They don't have to like you, and you don't hav...	0.0	0.0	0.0	0.0	NaN	Quality	NaN
14896	#Miami Graham Nash Live at Parker Playhouse #...	5647.0	15091.0	5823.0	0.0	United States	Spam	NaN
14897	@bethannhamilton is in the business of one-upp...	0.0	0.0	NaN	0.0	Southgate, MI	Quality	NaN
14898	Chasing Success by Space Cadetz Listen up...	1219.0	957.0	4077.0	1.0	United States	Spam	NaN

14899 rows x 8 columns



Limpeza dos dados

- Substituímos Nans pela mediana de cada coluna
- Removemos a classe 'South Dakota'
- Removemos as colunas localização e 'Unnamed 7'

```
=====
Tweet      0
following   158
followers   17
actions     3436
is_retweet   1
Type        0
dtypes: int64
```



Extração de Características

- A partir do texto do tweet calculamos 7 novas características.
- De 6 fomos para 14 características.

	Tweet	following	followers	actions	is_retweet	words	hashtags	hashtag_ratio	URLs	URL_ratio	numbers	mentions	Type
0	Good Morning Love @LeeBrown_V	0.0000	0.0000	0.0000	0.0000	5	0	0.0000	0	0.0000	0	1	Quality
1	'@realDonaldTrump @USNavy RIP TO HEROES'	42096.0000	61060.0000	5001.0000	0.0000	5	0	0.0000	0	0.0000	0	2	Spam
2	Haven't been following the news but I understa...	0.0000	0.0000	1496.0000	0.0000	14	1	0.0714	0	0.0000	0	0	Quality
3	pic.twitter.com/dy9q4fLhZ What to do with pap...	0.0000	0.0000	0.0000	0.0000	12	2	0.1667	2	0.1667	2	0	Quality
4	#DidYouKnow ► Mahatma Gandhi made a brief visi...	17800.0000	35100.0000	1496.0000	0.0000	17	2	0.1176	0	0.0000	2	1	Quality

Preparação dos Modelos

Tabela 3. Valores utilizados no Grid Search para Random Forest

Parâmetro	Valores	Descrição
<i>min_samples_split</i>	3, 5, 10	Número mínimo de amostras para considerar uma separação
<i>n_estimators</i>	100, 300	Total de árvores na floresta
<i>max_depth</i>	15, 25	Profundidade máxima de cada árvore
<i>max_features</i>	3, 6, 11	Características ao considerar a melhor separação

Tabela 4. Valores utilizados no Grid Search para KNN

Parâmetro	Valores	Descrição
<i>n_neighbors</i>	3, 5, 7, 10	Total de vizinhos a ser considerado
<i>algorithm</i>	auto, <i>ball_tree</i> , <i>kd_tree</i>	Algoritmo para o calculo de vizinhos
<i>metric</i>	euclidean, manhattan, minkowski	Métrica para o cálculo da distância

Tabela 5. Valores utilizados no Grid Search para MLP

Parâmetro	Valores	Descrição
<i>hidden_layer_sizes</i>	80, 100, 130	Número de neurônios nas camadas escondidas
<i>learning_rate_init</i>	0.001, 0.0001	Taxa de aprendizado inicial e de atualização
<i>max_iter</i>	100, 150, 200	Máximo de iterações



Preparação dos Modelos

- Random Forest
 - `n_estimators = 100, min_samples_split = 3, max_features = 3, max_depth = 15`
- KNN
 - `n_neighbors = 10, metric= manhattan, algorithm = auto`
- MLP
 - `hidden_layer_sizes = 80, learning_rate_init = 0.001, max_iter = 150`



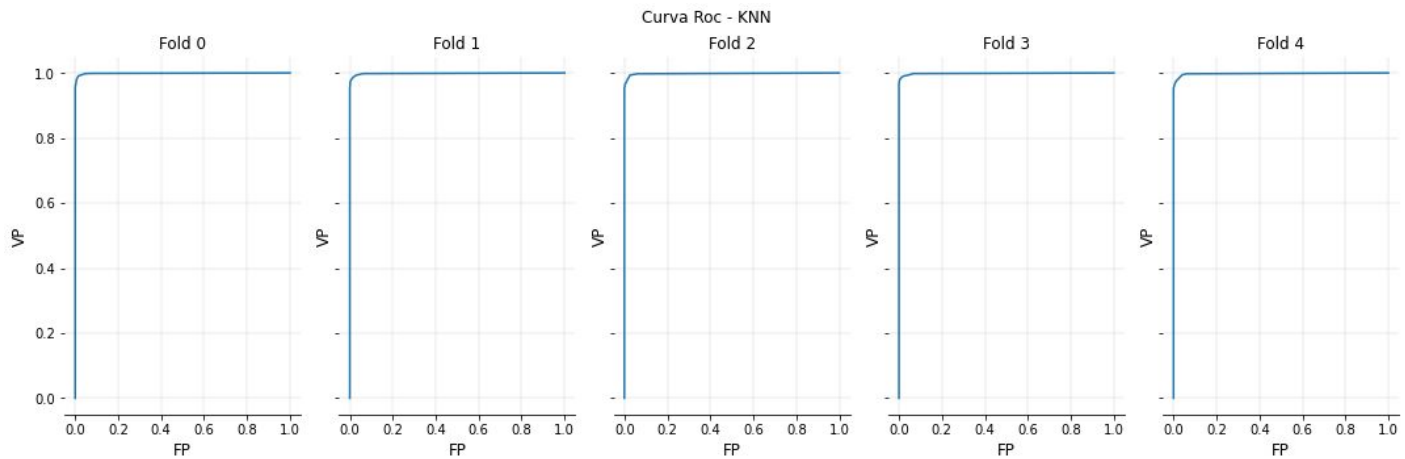
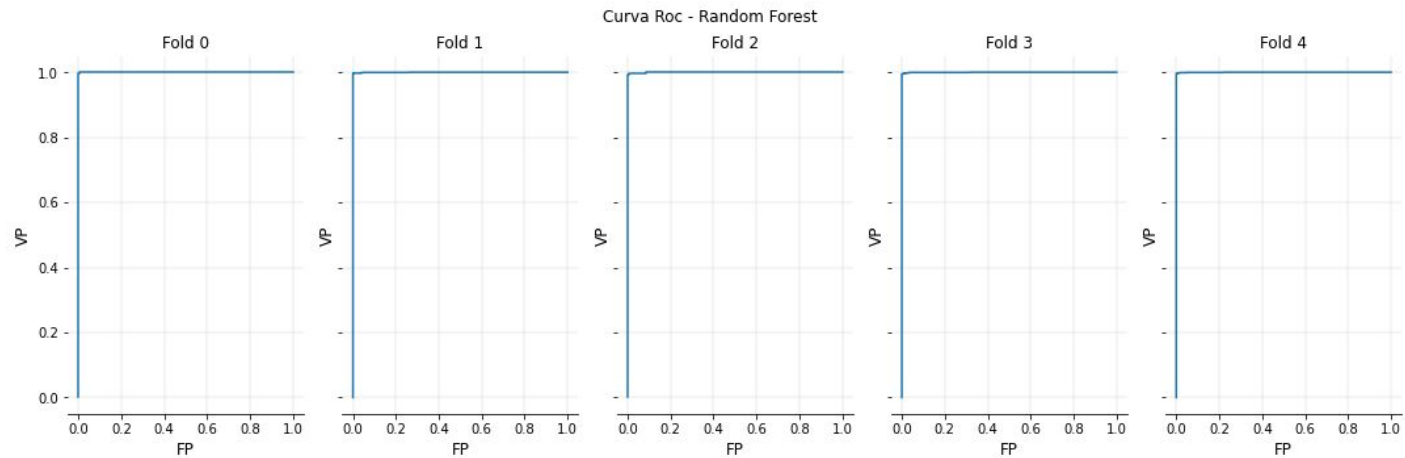
K-Fold Estratificado

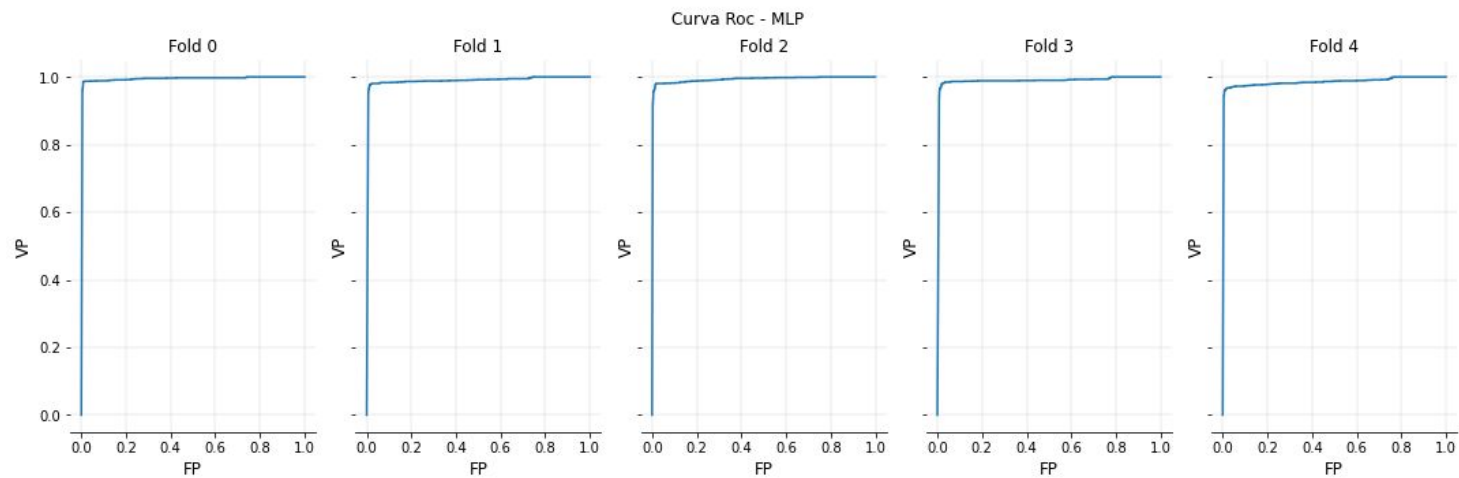
- Aplicamos o algoritmo de validação cruzada K-Fold estratificado com $k = 5$.
- Cada fold conta com 14.897 amostras sendo 11.918 para treinamento e 2.979 para teste.

Avaliação dos classificadores

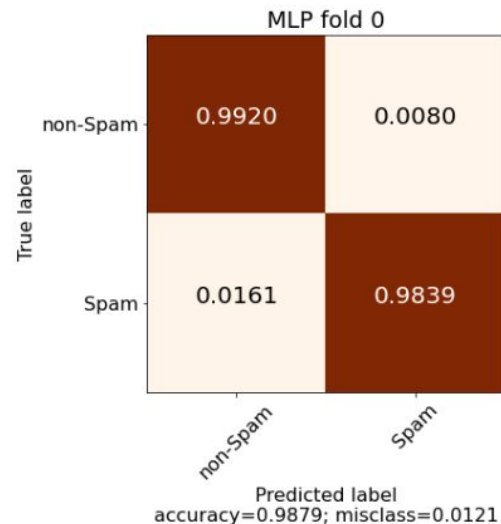
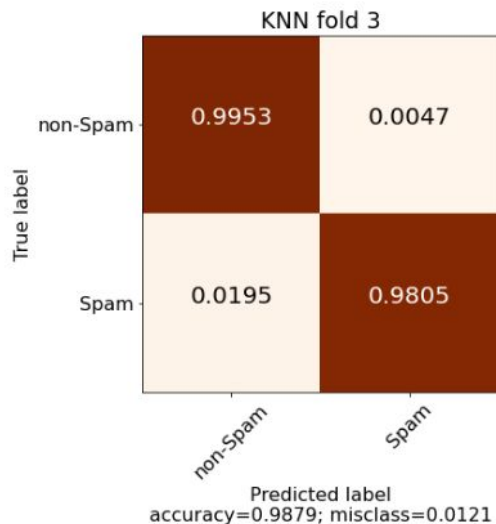
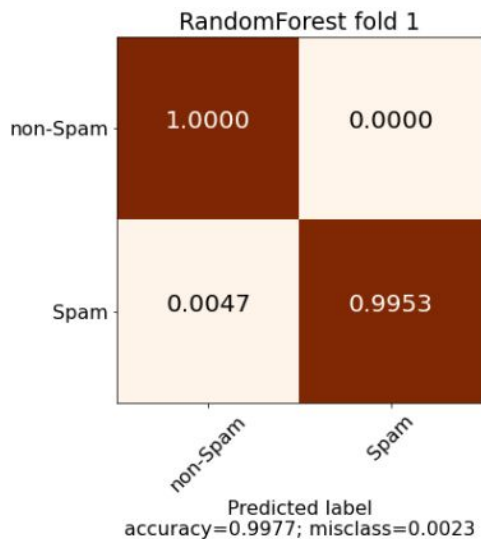
Tabela 6. Avaliação dos classificadores para cada porção de teste

		Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
Random Forest						
	Acurácia	1.00	1.00	0.99	1.00	1.00
	Precisão	1.00	1.00	1.00	1.00	1.00
	Recall	1.00	1.00	0.99	0.99	0.99
	F1 Score	1.00	1.00	0.99	1.00	1.00
KNN						
	Acurácia	0.99	0.99	0.98	0.99	0.98
	Precisão	0.99	0.99	0.99	1.00	0.99
	Recall	0.98	0.98	0.97	0.98	0.97
	F1 Score	0.99	0.99	0.98	0.99	0.98
MLP						
	Acurácia	0.99	0.98	0.98	0.98	0.98
	Precisão	0.99	0.99	0.98	0.98	0.99
	Recall	0.98	0.97	0.98	0.98	0.96
	F1 Score	0.99	0.98	0.98	0.98	0.98





Avaliação dos classificadores



Experimento - Política

	Tweet	following	followers	actions	is_retweet	words	hashtags	hashtag_ratio	URLs	URL_ratio	numbers	mentions
0	Primeira Dama Michele Bolsonaro agradecendo a ...	3165	5076	302	True	28	0	0.0000	1	0.0357	3	0
1	@pattisonbf parece os filhos do bolsonaro	107	529	0	False	6	0	0.0000	0	0.0000	0	1
2	@folha Quem tentou destruir a democracia foi o...	4	1	0	False	40	0	0.0000	0	0.0000	0	1
3	@TerraBrasilnot Acho q o imorais precisar ler ...	3975	6606	2	True	31	0	0.0000	0	0.0000	0	1
4	Rachadinha: suposto ex-funcionário de Carlos B...	2139	566	53	True	15	0	0.0000	1	0.0667	0	0
...
2477	Pela fase de grupos da UEFA Champions League, ...	6	28	0	False	26	0	0.0000	0	0.0000	5	0
2478	Lula é o maior ladrão de todos os tempos...#Lu...	357	471	7	True	10	1	0.1000	1	0.1000	3	0
2479	@pinto_saullo12 Se queremos derrotar Bolsonaro...	17783	16758	0	False	30	0	0.0000	1	0.0333	1	1
2480	@TradutordoBR Só observar os que odeiam BOLSON...	206	107	0	False	10	0	0.0000	0	0.0000	1	1
2481	CorreçãoCompartilhei uma FAKENEWS do Estadão, ...	1556	1197	675	True	38	0	0.0000	1	0.0263	3	0

2482 rows × 13 columns



Experimento - Política

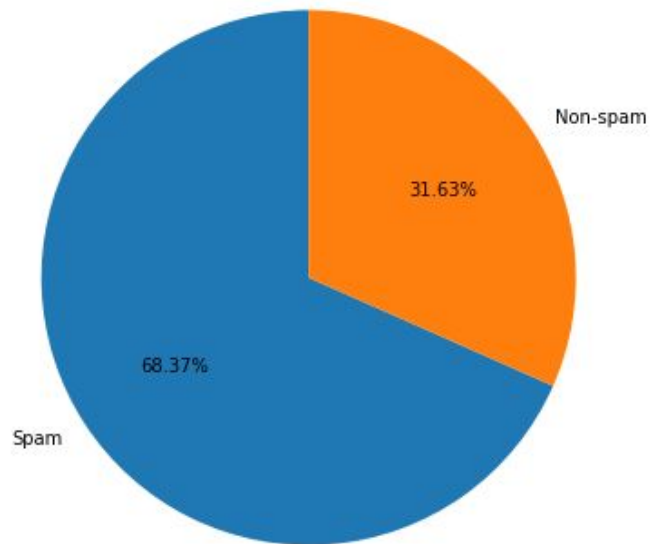
Características iniciais

	spam	non-spam
Tweets identificados	1697 (68.37%)	785 (31.63%)
Tweets únicos (sem repetição)	616	775
Tweets únicos/identificados	36.30%	98.73%
Total de urls	990	241
urls/identificados	58.34%	30.70%
mais de 1 url	62	11
mais de 1 url/total de urls	6.26%	4.56%

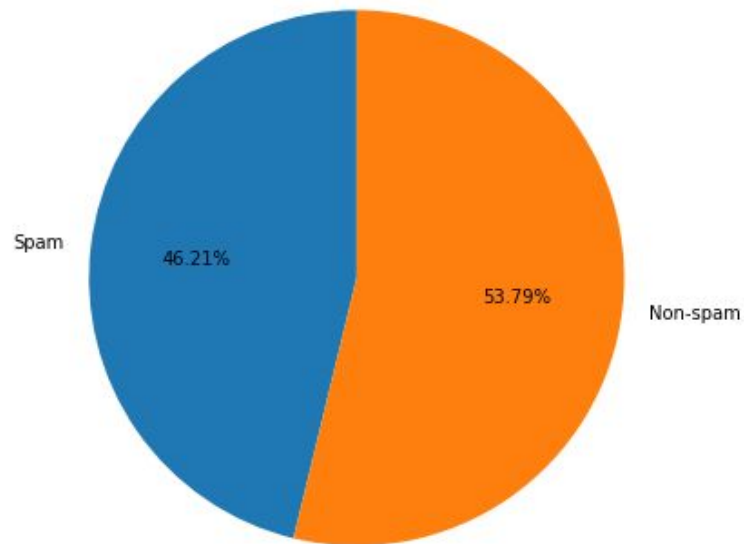
Experimento - Política

Relação entre spam e não spam

Relação entre todos os tweets

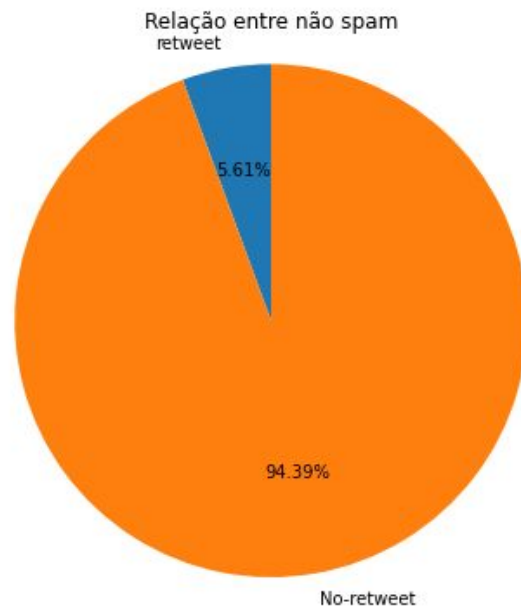
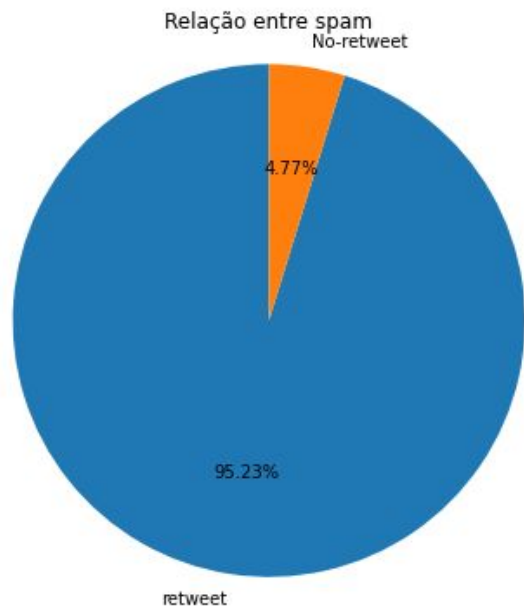


Relação entre tweets únicos (sem duplicados)



Experimento - Política

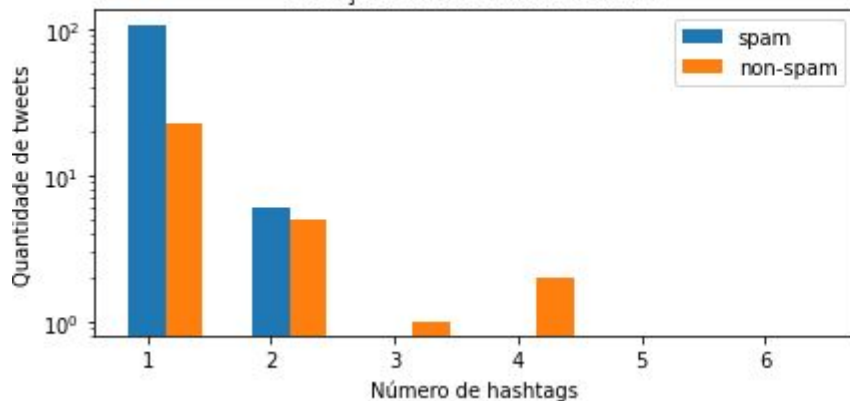
Relação entre retweet e não retweet



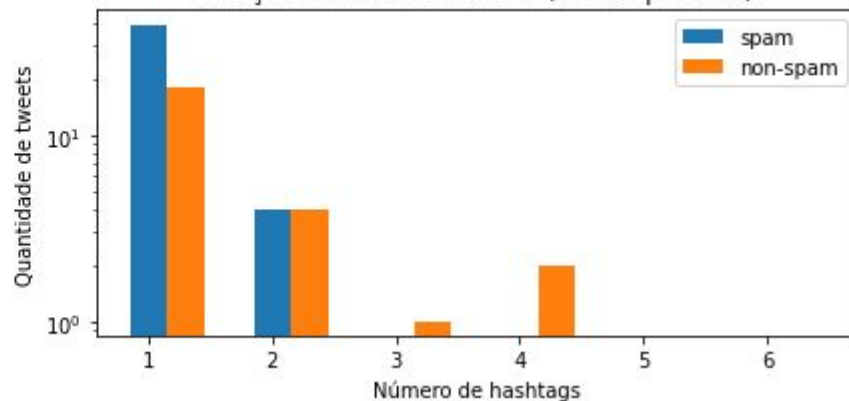
Experimento - Política

Quantidade de hashtags entre todos os tweets e tweets únicos

Relação entre todos os tweets



Relação entre tweets únicos (sem duplicados)





Experimento - Política

Top 5 Hashtags in spam and non-spam tweets
All tweets:

	Hashtags	Quantidade	Hashtags	Quantidade
0	#MissGrandInternational	45	#ForaBolsonaro	4
1	#BolsonaroNuncaMais	7	#PTNuncaMais	2
2	#LulaEoPTMulheresComDireitos	7	#GrandLorena	2
3	#ForaBolsonaro	6	#STF	2
4	#tacladuran	5	#LulaLadrao	2



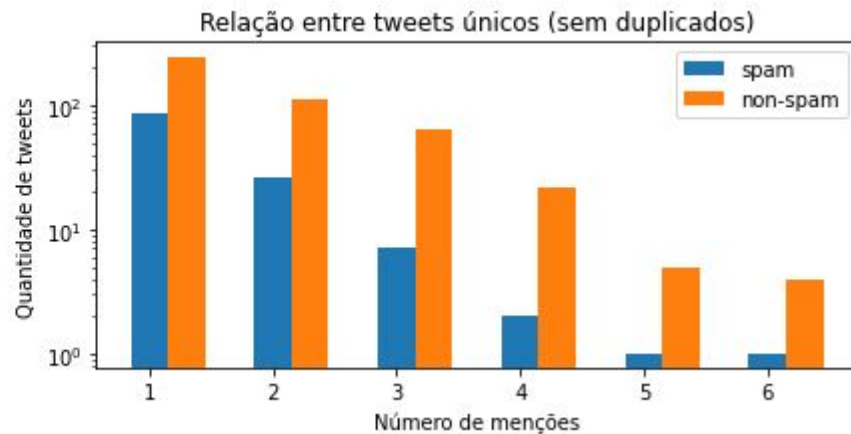
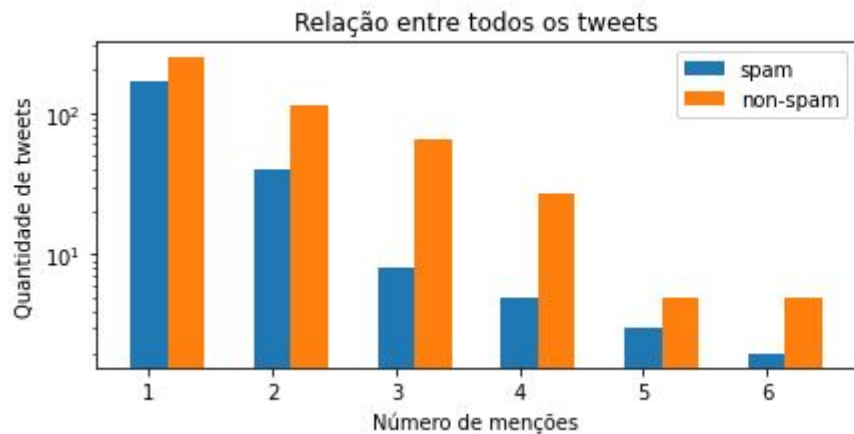
Experimento - Política

Top 5 Hashtags in spam and non-spam tweets
All tweets:

	Hashtags	Quantidade	Hashtags	Quantidade
0	#MissGrandInternational	6	#ForaBolsonaro	3
1	#ForaBolsonaro	4	#PTNuncaMais	2
2	#BolsonaroNuncaMais	3	#STF	2
3	#LulaEoPTMulheresComDireitos	3	#LulaLadrao	2
4	#LulaNoPodpah	2	#MoroTraidor	1

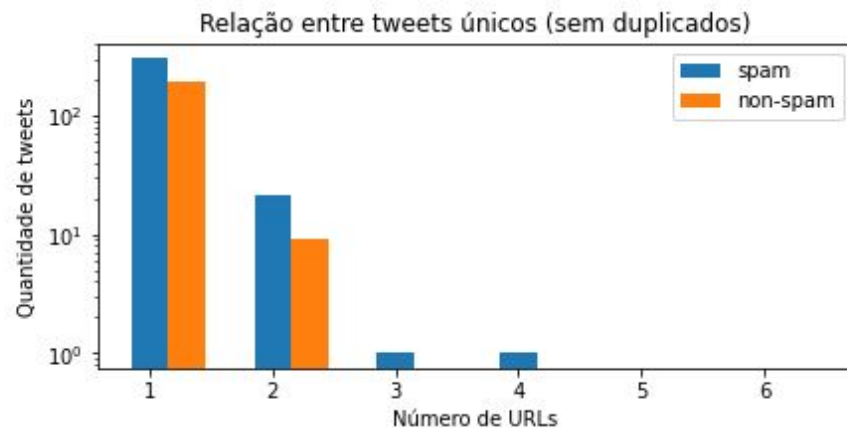
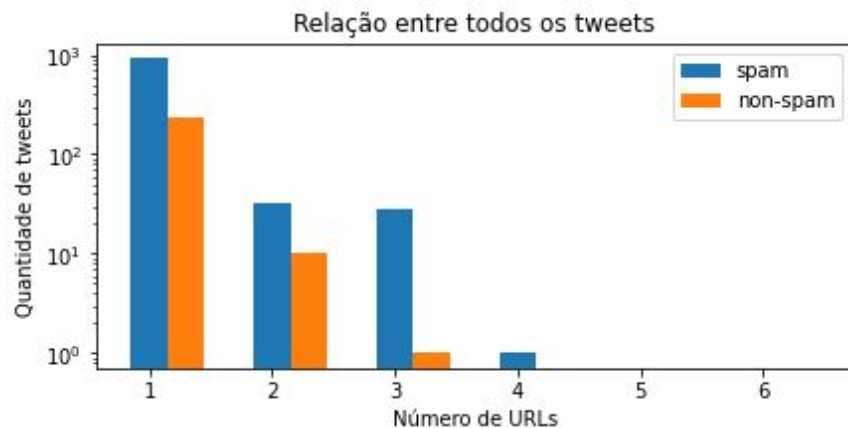
Experimento - Política

Quantidade de menções entre todos os tweets e tweets únicos



Experimento - Política

Quantidade de URLs entre todos os tweets e tweets únicos





Experimento - Política

Top 5 URLs in spam and non-spam tweets
All tweets:

URLs		Quantidade	URLs		Quantidade
0	https://twitter.com/Jouberth19/status/14671085...	48	https://www.youtube.com/channel/UCGfNDjbkEK89D...		16
1	https://twitter.com/PATRIOTAS/status/146714514...	48	https://twitter.com/Simonycorredora/status/146...		4
2	https://uc604efede1f22fd46e14600630c.dl.dropbo...	28	https://twitter.com/phvitoria1/status/14553239...		3
3	https://t.me/desmentindobofo	28	https://pleno.news/brasil/eleicoes-2022/partid...		2
4	https://twitter.com/desmentindobofo/status/146...	28	https://twitter.com/Simonycorredora/status/146...		2



Experimento - Política

Top 5 URLs in spam and non-spam tweets
All tweets:

URLs		Quantidade	URLs		Quantidade
0	https://twitter.com/Jouberth19/status/14671085...	1	https://www.youtube.com/channel/UCGfNDjbkEK89D...		16
1	https://www.brasil247.com/regionais/sudeste/ra...	1	https://twitter.com/phvitoria1/status/14553239...		2
2	https://revistaoeste.com/politica/partido-de-b...	1	https://twitter.com/Marcia06513945/status/1467...		1
3	https://revistaoeste.com/politica/bolsonaro-af...	1	https://www1.folha.uol.com.br/equilibrioesaude...		1
4	https://terrabrasilnoticias.com/2021/12/michel...	1	https://www.youtube.com/watch?v=7wNCmZuw0aw&fe...		1

Experimento - Sorteio

	Tweet	following	followers	actions	is_retweet	words	hashtags	hashtag_ratio	URLs	URL_ratio	numbers	mentions
0	Fui às minhas DMs e o @hearthstone_pt disse qu...	1605	141	1157	True	40	0	0.0000	1	0.0250	6	2
1	O Natal chega mais cedo com a @ftwesports @OME...	35	2	28	True	28	2	0.0714	2	0.0714	3	2
2	@TWICE_Giveaway @JYPETWICE Sana espresso#SCIEN...	451	368	0	False	8	1	0.1250	1	0.1250	1	3
3	@TWICE_Giveaway @JYPETWICE Momo espresso#SCIEN...	451	368	0	False	8	1	0.1250	1	0.1250	2	3
4	Essa dupla é incrível demais para ficar conosc...	2	0	132	True	42	0	0.0000	2	0.0476	5	0
...
2493	Fui às minhas DMs e o @hearthstone_pt disse qu...	1001	39	1157	True	40	0	0.0000	1	0.0250	6	2
2494	Fui às minhas DMs e o @hearthstone_pt disse qu...	305	14	1157	True	40	0	0.0000	1	0.0250	6	2
2495	Fui às minhas DMs e o @hearthstone_pt disse qu...	990	41	1157	True	40	0	0.0000	1	0.0250	6	2
2496	SORTEIO 🎁 Para comemorar os 5000 inscritos no ca...	314	35	1021	True	33	2	0.0606	1	0.0303	6	0
2497	Monstruoso este giveaway de natal da @ftwespor...	103	442	40	True	31	4	0.1290	2	0.0645	5	2

2498 rows × 13 columns



Experimento - Sorteio

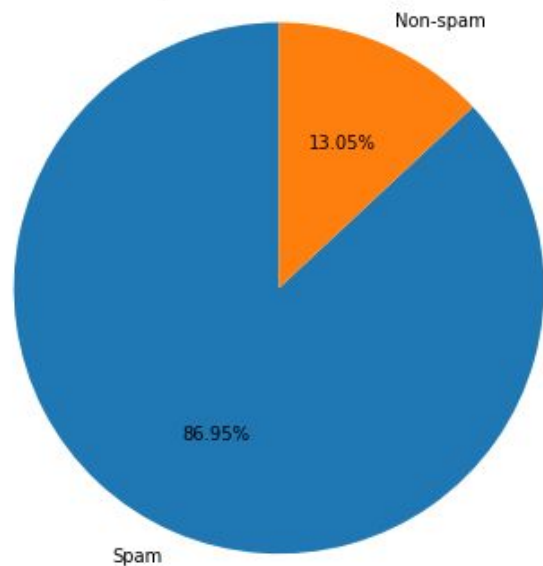
Características iniciais

	spam	non-spam
Tweets identificados	2172 (86.95%)	326 (13.05%)
Tweets únicos (sem repetição)	94	235
Tweets únicos/identificados	4.33%	72.09%
Total de urls	2131	191
urls/identificados	98.11%	58.59%
mais de 1 url	330	13
mais de 1 url/total de urls	15.49%	6.81%

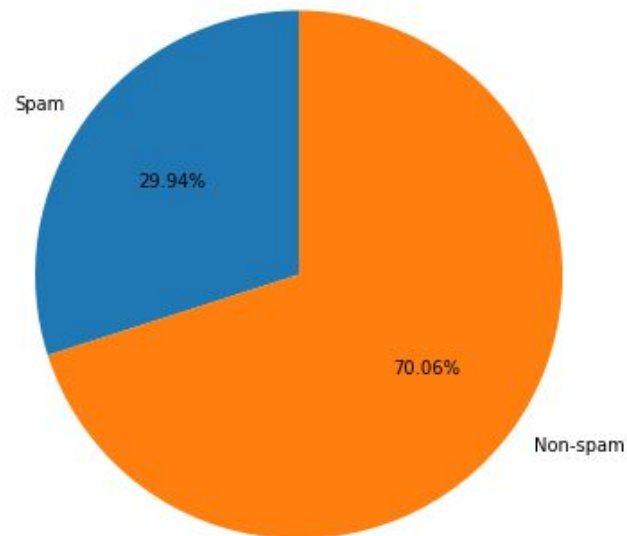
Experimento - Sorteio

Relação entre spam e não spam

Relação entre todos os tweets

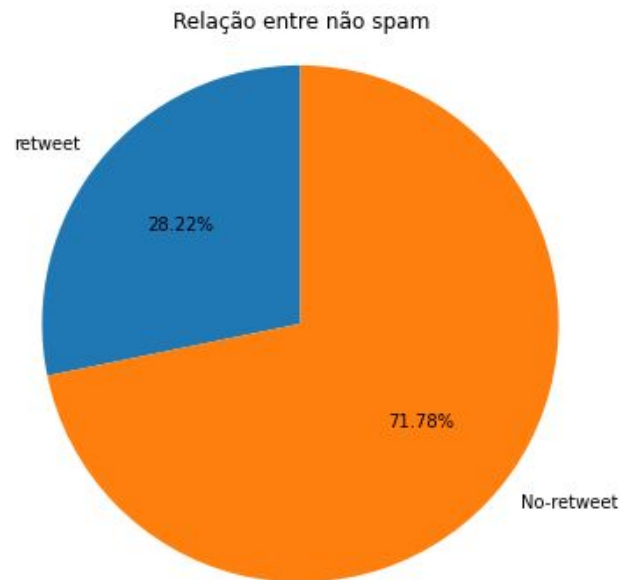
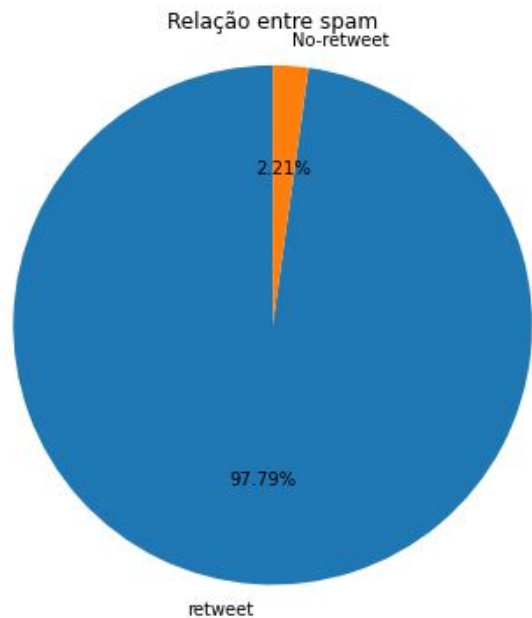


Relação entre tweets únicos (sem duplicados)



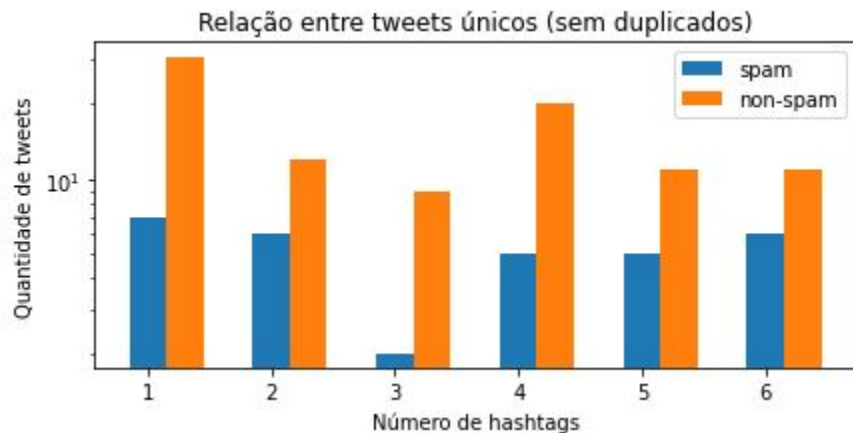
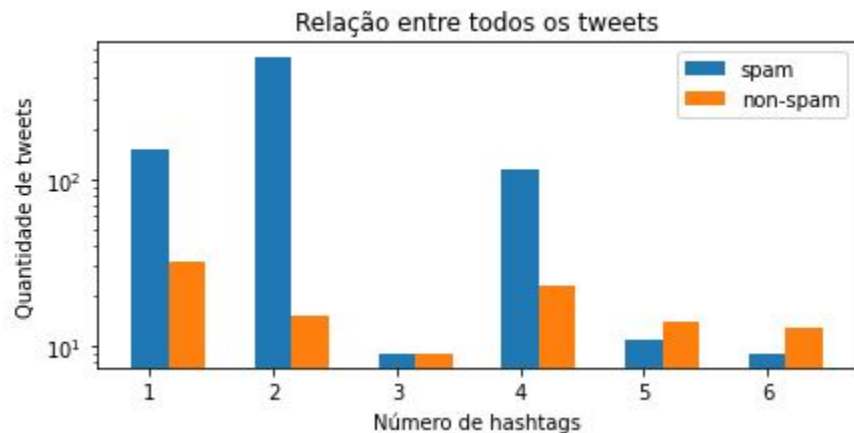
Experimento - Sorteio

Relação entre retweet e não retweet



Experimento - Sorteio

Quantidade de hashtags entre todos os tweets e tweets únicos





Experimento - Sorteio

Top 5 Hashtags in spam and non-spam tweets
All tweets:

	Hashtags	Quantidade	Hashtags	Quantidade
0	#giveaway	570	#Giveaway	65
1	#Hearthstone	508	#giveaway	28
2	#OMENFTW	131	#NFTs	21
3	#parceria	131	#Airdrop	20
4	#ftwesports	119	#NFTGiveaway	20



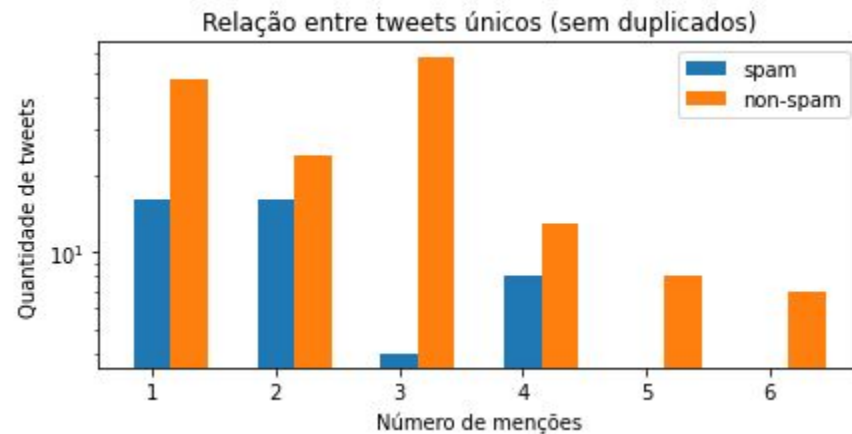
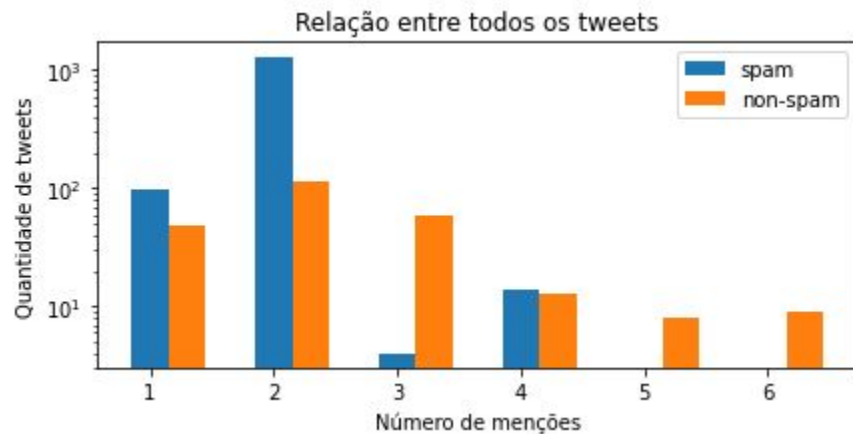
Experimento - Sorteio

Top 5 Hashtags in spam and non-spam tweets
All tweets:

	Hashtags	Quantidade	Hashtags	Quantidade
0	#Giveaway	23	#Giveaway	59
1	#giveaway	12	#giveaway	22
2	#NFTGiveaway	8	#NFTs	21
3	#Sorteio	6	#Airdrop	20
4	#xboxonex	6	#NFTGiveaway	20

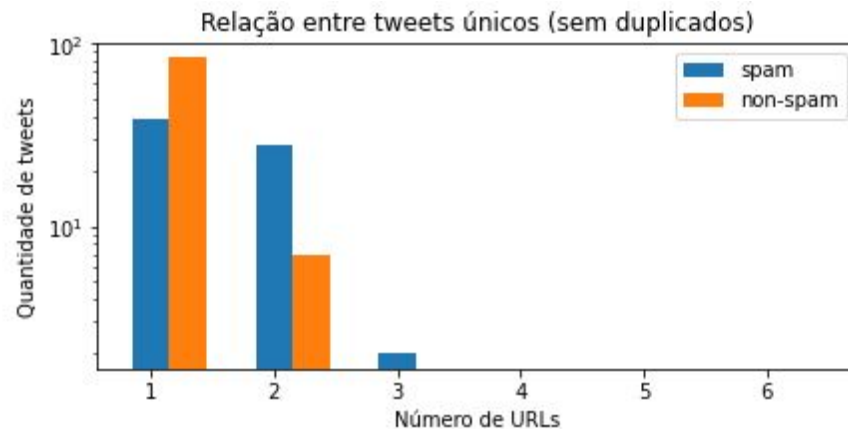
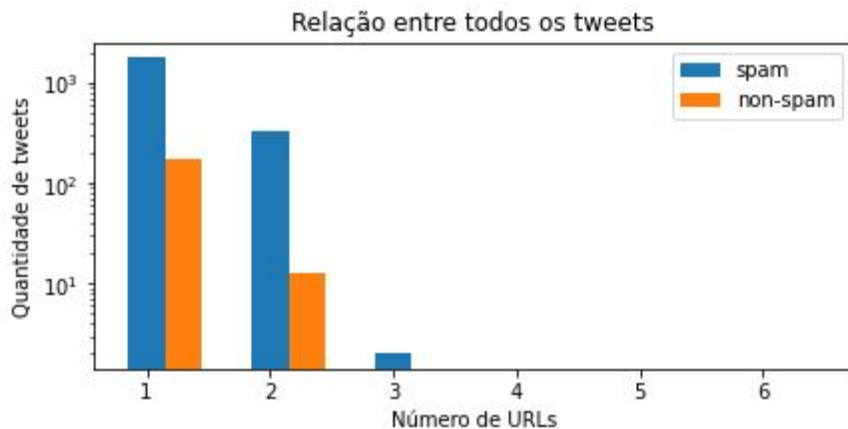
Experimento - Sorteio

Quantidade de menções entre todos os tweets e tweets únicos



Experimento - Sorteio

Quantidade de URLs entre todos os tweets e tweets únicos





Experimento - Sorteio

Top 5 URLs in spam and non-spam tweets
All tweets:

URLs Quantidade			URLs Quantidade	
0	https://twitter.com/JOliveira10_/status/146673...	567	https://twitter.com/SuperFakeHS/status/1466011...	86
1	https://twitter.com/KingVenomStream/status/146...	508	https://twitter.com/KingVenomStream/status/146...	3
2	https://twitter.com/SuperFakeHS/status/1466011...	494	https://www.gleamsorteios.com/2021/12/sorteio-...	2
3	https://gleam.io/hZOOZ/logitech-g-x-astro-2021...	122	https://twitter.com/GleamSorteios/status/14665...	2
4	https://twitter.com/ASTROGamingBR/status/14664...	122	https://twitter.com/mihyunvibe/status/14671390...	1



Experimento - Sorteio

Top 5 URLs in spam and non-spam tweets
All tweets:

URLs		Quantidade	URLs		Quantidade
0	https://botisimo.com/u/tiorapadura/shop	3	https://twitter.com/mihyunvibe/status/14671390...		1
1	https://gleam.io/3yLbY/sorteo-ps5-4-juegos?gsr...	2	https://twitter.com/mihyunvibe/status/14671389...		1
2	https://twitter.com/SuperFakeHS/status/1466011...	1	https://twitter.com/itsnrjonce/status/14671366...		1
3	https://gleam.io/cjdEt/ftw-omen-squad-giveaway...	1	https://twitter.com/ilikeSweetberry/status/146...		1
4	https://twitter.com/MoraisGaming/status/146613...	1	https://twitter.com/seokjinportugal/status/145...		1