

Twitter e Eleições

Palanque de Bolso(naro)

Qual a correlação dos tweets de bolsonaro com sua popularidade como candidato?

Parte I

Os tweets

	created_at	text	retweet_count	favorite_count
0	2018-10-28 23:28:12	Jair Bolsonaro fala pela primeira vez como Pre...	5199	33917
1	2018-10-28 23:26:01	Obrigado pela confiança! Vamos juntos mudar o ...	29207	135263
2	2018-10-28 23:13:01	BRASIL! BR👍	51825	211283
3	2018-10-28 01:38:42	@lobaoeletrico 👍 BR	565	4489
4	2018-10-28 00:59:06	ATÉ A VITÓRIA, SE DEUS QUISER! BRBRBRBR https:...	19733	91936



Mineração do texto

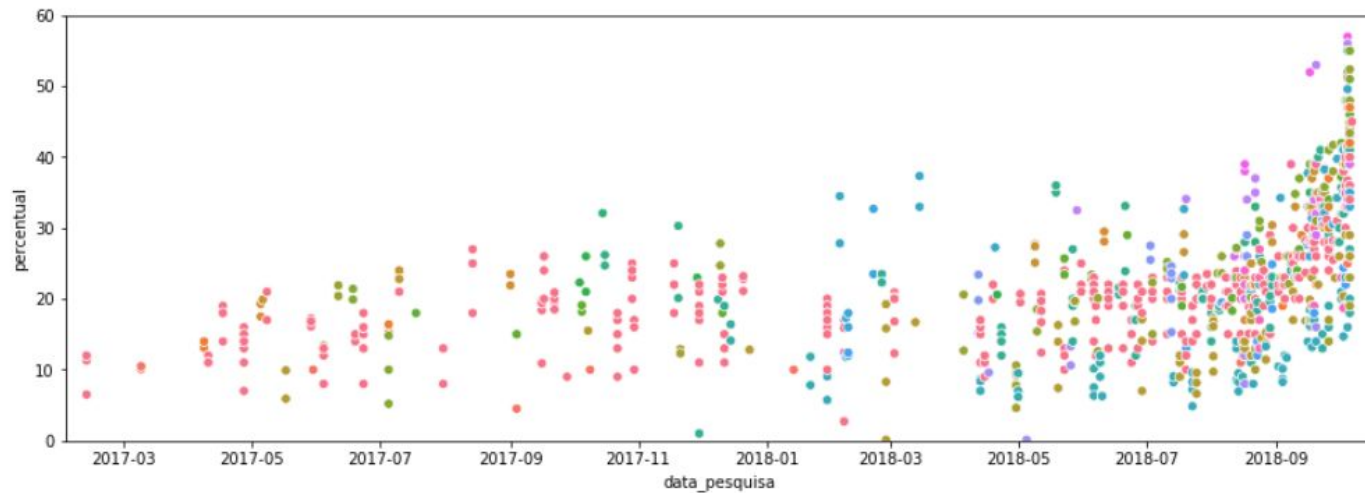
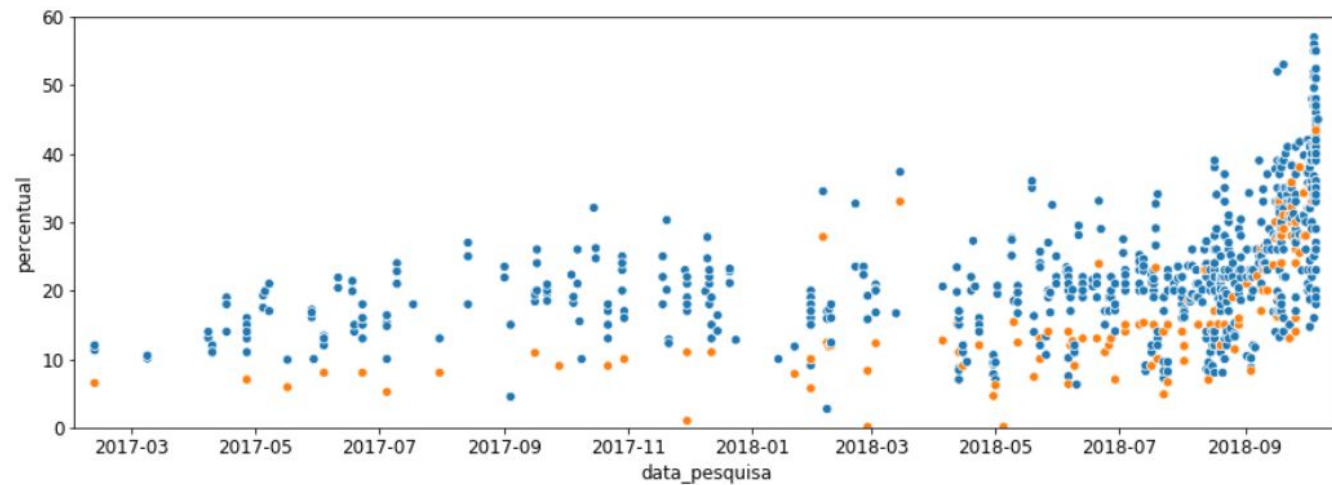


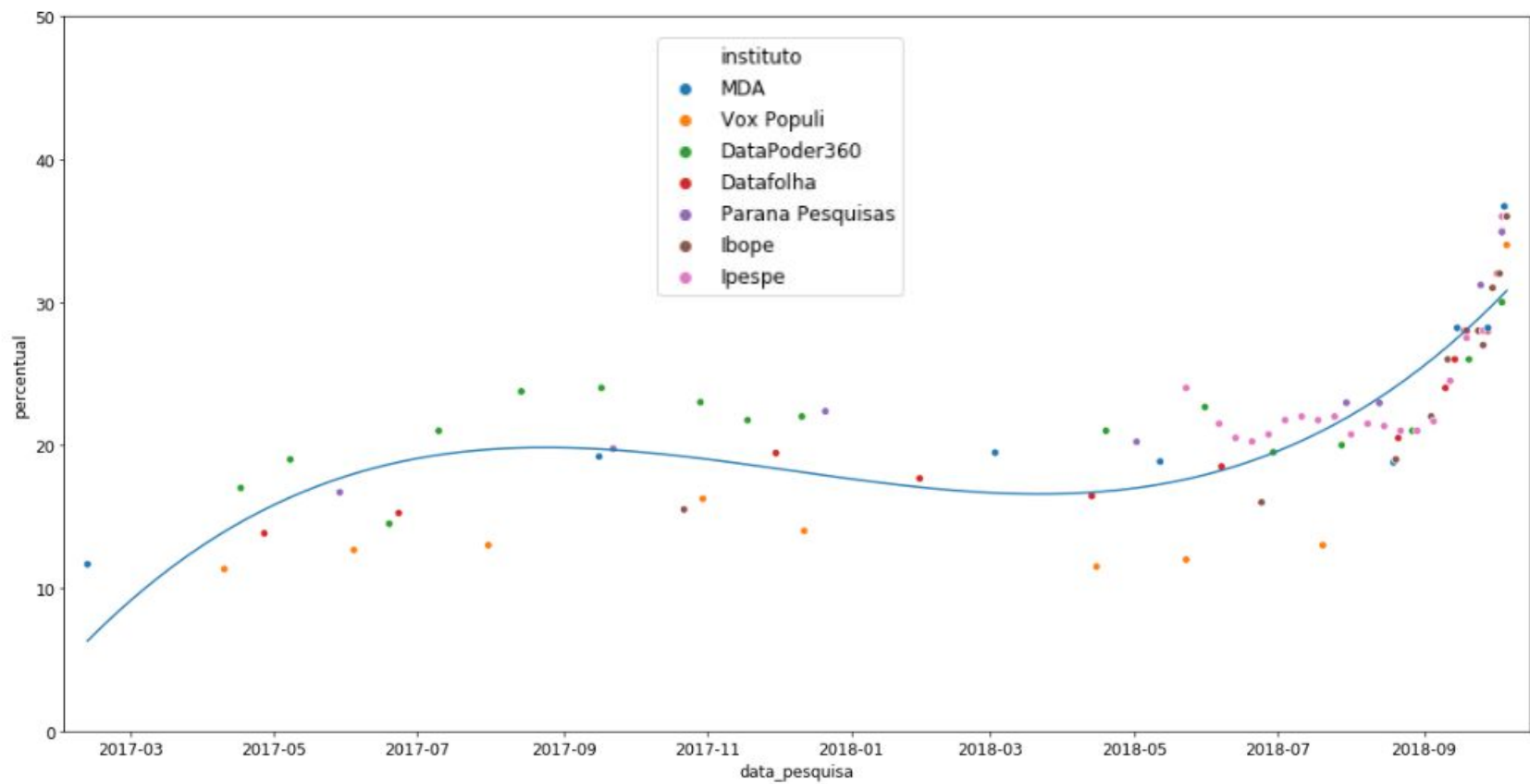
	sentimento	subjetividade	ADJ	ADP	ADV	CONJ	DET	NOUN	NUM	PRT	PRON	VERB	.	X
0	0.25	0.333333	2	2	0	0	1	5	0	0	0	1	0	0
1	0.00	0.000000	0	1	1	0	0	5	0	1	2	2	0	0
2	0.00	0.000000	0	0	0	0	0	2	0	0	0	0	0	0
3	0.00	0.000000	1	0	0	0	0	2	0	0	0	0	0	0
4	1.00	0.400000	1	0	0	0	0	6	0	0	0	1	0	0

SEGUNDA Parte ou Parte B

Classe alvo: Popularidade

	pesquisa_id	instituto	unidade_federativa_nome	tipo	voto_tipo	data_pesquisa	partido	percentual
0	3517	MDA	BRASIL	Estimulada	Votos Totais	2017-02-11	PSC	11.7
1	3517	MDA	BRASIL	Estimulada	Votos Totais	2017-02-11	PSC	11.3
2	3517	MDA	BRASIL	Estimulada	Votos Totais	2017-02-11	PSC	12.0
3	3517	MDA	BRASIL	Espontanea	Votos Totais	2017-02-11	PSC	6.5
4	7630	Parana Pesquisas	ALAGOAS	Estimulada	Votos Totais	2017-03-09	PSL	10.2
5	7630	Parana Pesquisas	ALAGOAS	Estimulada	Votos Totais	2017-03-09	PSL	10.1





Parte III

Tá... e os resultados?

Sem mineração de dados

```
regr = RandomForestRegressor(n_jobs=-1,  
                             random_state=42)  
regr.fit(x_train, y_train)
```

```
regr.score(x_train, y_train), regr.score(x_test, y_test)
```

(0.34258292360783504, -0.1425513458351877)

r2 negativo... que complexo

Com mineração de dados

```
regr = RandomForestRegressor(n_jobs=-1,  
                             random_state=42)  
regr.fit(x_train, y_train)
```

```
regr.score(x_train, y_train), regr.score(x_test, y_test)
```

(0.9325329111353104, 0.6074960974619352)

Melhorias

```
RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=15,  
                        max_features='auto', max_leaf_nodes=None,  
                        min_impurity_decrease=0.0, min_impurity_split=None,  
                        min_samples_leaf=1, min_samples_split=5,  
                        min_weight_fraction_leaf=0.0, n_estimators=400, n_jobs=-1,  
                        oob_score=False, random_state=42, verbose=0,  
                        warm_start=False)
```

(0.9250217111181774, 0.7085966677521647)

- Gridsearch
- Mais métricas de sentimento
- Normalização dos dados
- Filtragens diferentes para gerar a classe alvo
- Impeachment
- Despolarização de bolhas sociais
- Participação política

Existem vários modelos de classificação de sentimento e polaridade. Num próximo passo, pretendemos usar esses outros métodos como alternativa ou em conjunto.

Não há necessidade da normalização dos dados na Floresta já que cada coluna é comparada individualmente, ao invés de interagir com dados de outras colunas.

Durante o trabalho, sentimos a enorme necessidade e vantagem da interdisciplinaridade quando estamos trabalhando com dados de áreas que não a nossa.

Considerações Finais

Sugestões, dúvidas?



Toda arrogância será
castigada...

That's all folks!

