

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
ESCOLA DE CIÊNCIAS & TECNOLOGIA

IDENTIFICADOR DE SENTIMENTO EM POSTS DO TWITTER

Docente: Orivaldo Vieira De Santana Junior
Discente: Alysson Rafael Oliveira de Lima

Proposta

O projeto busca desenvolver uma ferramenta para identificar sentimentos em posts do twitter sendo considerado 3 classes: **Positivos**, **Negativos** e **Neutros**.

Modelos Adotados - NLP

Para identificar os sentimentos utilizaremos o processamento de linguagem natural (NLP) que trata do desenvolvimento de aplicativos e serviços capazes de entender idiomas humanos.

Modelos Adotados - SOM

Após os dados serem categorizados sentimentalmente, serão aplicados a um modelo de rede neural SOM, para ser extraído as características e ser possível algum agrupamento automático.

Divisão

O desenvolvimento foi dividido em três partes:

- ❖ Mineração dos dados.
- ❖ Aplicação do processamento de linguagem Natural.
- ❖ Aplicação da rede SOM.

Obs.:

O algoritmo foi desenvolvido para o idioma Inglês, visto que a maioria das funcionalidades estão configuradas assim. Não havendo tempo hábil para mudança.

Mineração dos dados

Foi utilizado a API do twitter, disponível na plataforma de desenvolvedor, para obter os tweets públicos dos perfis.

Tweepy

Para facilitar o trabalho com a API foi utilizado a biblioteca para python Tweepy.

Tweepy

#conecta na conta e exibe os tweets públicos

```
import tweepy
```

```
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)  
auth.set_access_token(access_token, access_token_secret)
```

```
api = tweepy.API(auth)
```

```
public_tweets = api.home_timeline()  
for tweet in public_tweets:  
    print(tweet.text)
```

Método utilizado para obter os dados

```
def obter_tweets(usuario, limite=500):  
    resultados = api.user_timeline(screen_name=usuario,  
count=limite, tweet_mode='extended')  
    tweets = []  
    for r in resultados:  
        tweet = re.sub(r'http\S+', '', r.full_text)  
        tweets.append(tweet.replace('\n', ' '))  
    return tweets
```

1		Tweets
2	0	95% Approval Rating in the Republican Party. Thank you!
3	1	New Stock Market Record today, AGAIN. Congratulations USA!
4	2	People now realize it is a Democrat Hoax!
5	3	RT @piersmorgan: 🤔🤔
6	4	RT @EricTrump: What an amazing night in Florida! #USAUSAUSA
7	6	RT @GOPChairwoman: Overwhelmingly strong support for @realDonaldTrump in a DEM county in battleground Florida. His supporters are already...
8	7	GOD BLESS THE U.S.A.! #MAGA
9	9	THANK YOU FLORIDA! #KAG2020
10	10	"OMB Official: Ukraine Aid Held Up Because Other Countries Weren't Giving"
11	12	RT @WhiteHouse: LIVE: President @realDonaldTrump Pardons the National Thanksgiving Turkey
12	13	...love to have Mike Pompeo, Rick Perry, Mick Mulvaney and many others testify about the phony Impeachment Hoax. It is a Democrat Scam that is going nowhere but, fu
13	14	...lawyer has already stated that I did nothing wrong. John Bolton is a patriot and may know that I held back the money from Ukraine because it is considered a corrupt co
14	15	The D.C. Wolves and Fake News Media are reading far too much into people being forced by Courts to testify before Congress. I am fighting for future Presidents and the
15	16	RT @DailyCaller: President @realDonaldTrump signs the Women's Suffrage Centennial Commemorative Coin Act which will direct the U.S. Treasur...

Aplicação do modelo NLP

Foi utilizado as seguintes configurações:

- ❖ Biblioteca: **NLTK**
- ❖ classificador: **Naive Bayes classifier**
- ❖ Analisador de sentimentos: **Vader Lexicon**

Resultado da análise do Vader

```
“VADER is smart, handsome, and funny.”
```

```
{'pos': 0.746, 'compound': 0.8316, 'neu': 0.254, 'neg': 0.0}
```

```
“VADER is smart, handsome, and funny!”
```

```
{'pos': 0.752, 'compound': 0.8439, 'neu': 0.248, 'neg': 0.0}
```

```
“VADER is VERY SMART, handsome, and FUNNY.”
```

```
{'pos': 0.754, 'compound': 0.9227, 'neu': 0.246, 'neg': 0.0}
```

Unnamed: 0		tweets	compound	negativos	neutro	positivo
0	Before arguing with friends or family around t...	0.6341	0.115	0.671	0.214	
1	Politicians shouldn't be picking their voters....	0.3182	0.000	0.944	0.056	
2	Take a look at this piece to understand the ve...	0.2500	0.000	0.950	0.050	
3	Young leaders like these are powering the prom...	0.8553	0.000	0.722	0.278	
4	Proud to see Oluwaseun @AyodejiOsowobi in this...	0.1779	0.087	0.789	0.124	
5	Stories like this are a reminder of our duty, ...	0.3818	0.079	0.754	0.167	
6	This Veterans Day, here's a moving portrait of...	0.7579	0.000	0.866	0.134	
7	No one says it better than @MichelleObama — th...	0.1779	0.069	0.841	0.090	
8	Proud of all the Americans who showed up to vo...	0.8271	0.145	0.565	0.290	
9	Proud to endorse an outstanding group of Virgi...	0.9684	0.000	0.603	0.397	

Aplicação do modelo SOM

```
#tamanho da rede
```

```
tx = 4
```

```
ty = 4
```

```
# Training the SOM
```

```
from minisom import MiniSom
```

```
som = MiniSom(x = tx, y = ty, input_len = 4, sigma = 1.0,
```

```
learning_rate = 0.5)
```

```
som.random_weights_init(X)
```

```
som.train_random(data = X, num_iteration = 1000)
```

Resultado

Para analisar o processo foi utilizado os perfis de dois presidentes dos Estados Unidos da América, Donald Trump e Barack Obama. Foram colhidos 200 tweets de cada um para análise.

Textos utilizados como parâmetro

HAPPY!!

compound: 0.6467, neg: 0.0, neu: 0.0, pos: 1.0,

BAD!!

compound: -0.6229, neg: 1.0, neu: 0.0, pos: 0.0,

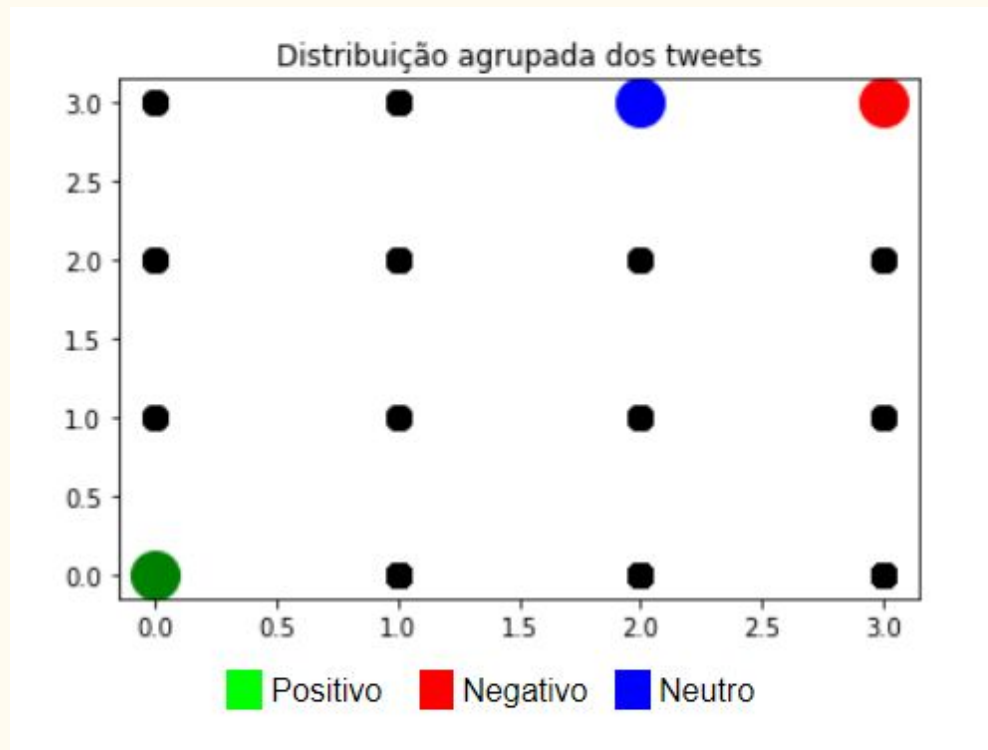
NOT

compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,

sky is blue!!

compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,

Estudo de caso - Barack Obama



Total Barack Obama:

`[[27. 7. 6. 6.]`

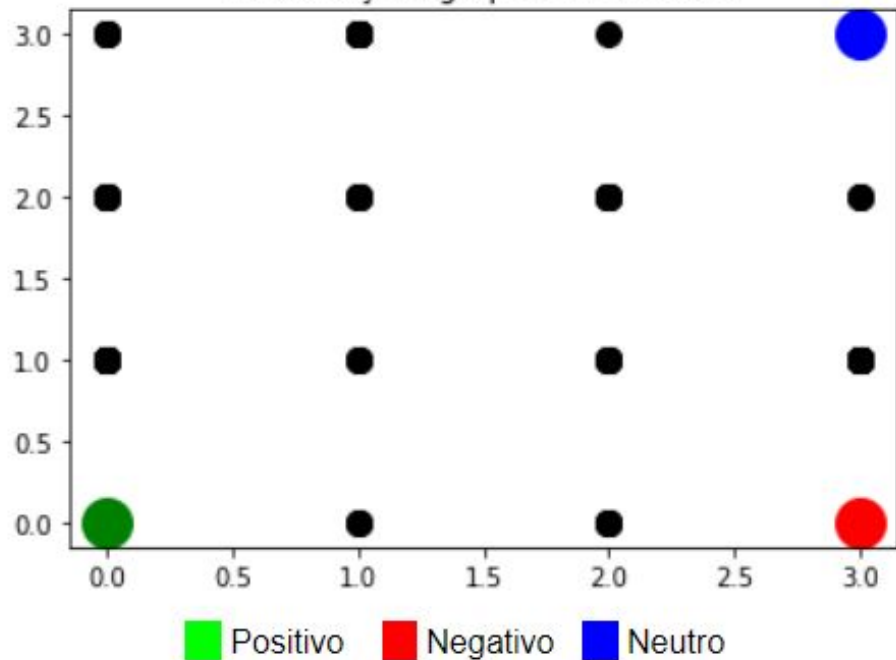
`[25. 7. 10. 11.]`

`[22. 10. 7. 16.]`

`[19. 9. 7. 10.]]`

Estudo de caso - Trump

Distribuição agrupada dos tweets



Total Donald Trump:

`[[17. 19. 10. 6.]`

`[3. 4. 7. 13.]`

`[5. 7. 8. 1.]`

`[19. 20. 3. 44.]]`

Conclusão

Através da combinação de NLP com o modelo SOM foi possível obter bons resultados no reconhecimento de padrões em textos. Com ajustes é possível algo mais específico, como reconhecimento de posts de ódios ou até mesmo tendências suicidas e etc.

FIM