



Departamento de Ciência da Computação
Universidade de Brasília

Análise sentimental textual de tweets em língua portuguesa sobre a pandemia de covid-19 no Brasil

Artur Filgueiras Scheiba Zorron

Orientador: Vinícius Pereira Gonçalves

24 de fevereiro de 2023

Agenda



- 01 Introdução
- 02 Trabalhos relacionados
- 03 Desenvolvimento
- 04 Resultados
- 05 Conclusão
- 06 Referências

- 01** | **Introdução**
- 02 | Trabalhos relacionados
- 03 | Desenvolvimento
- 04 | Resultados
- 05 | Conclusão
- 06 | Referências

- Com o início da pandemia de covid-19, as pessoas começaram a se comunicar cada vez mais pelas redes sociais.
- Uma das redes sociais mais utilizadas foi o Twitter, rede social baseada em textos curtos e de consumo rápido, também conhecidos como tweets.

- Este trabalho possui o objetivo de entender como os brasileiros se expressaram sentimentalmente no Twitter com base em seus tweets.
- Para isso, foi feita uma análise sentimental textual em tweets publicados em língua portuguesa sobre a pandemia de covid-19 no Brasil.

- A hipótese inicial é uma análise sentimental com uma maior quantidade de tweets de cunho negativo, pois nesse período muitas pessoas passaram por situações ruins como perdas de entes queridos, etc.

Agenda



- 01 Introdução
- 02 Trabalhos relacionados**
- 03 Desenvolvimento
- 04 Resultados
- 05 Conclusão
- 06 Referências

Trabalhos relacionados: Comparação

Autores	Covid	Twitter	Análise Sentimental	BERT	Brasil	Língua Portuguesa
Mansoor et al.	X	X	X		X	
Kausar et al.	X	X	X		X	
Melo et al.	X	X	X		X	X
Sitaula et al.	X	X	X	X		
Liu et al.	X	X	X	X		
Zorron et al.	X	X	X	X	X	X

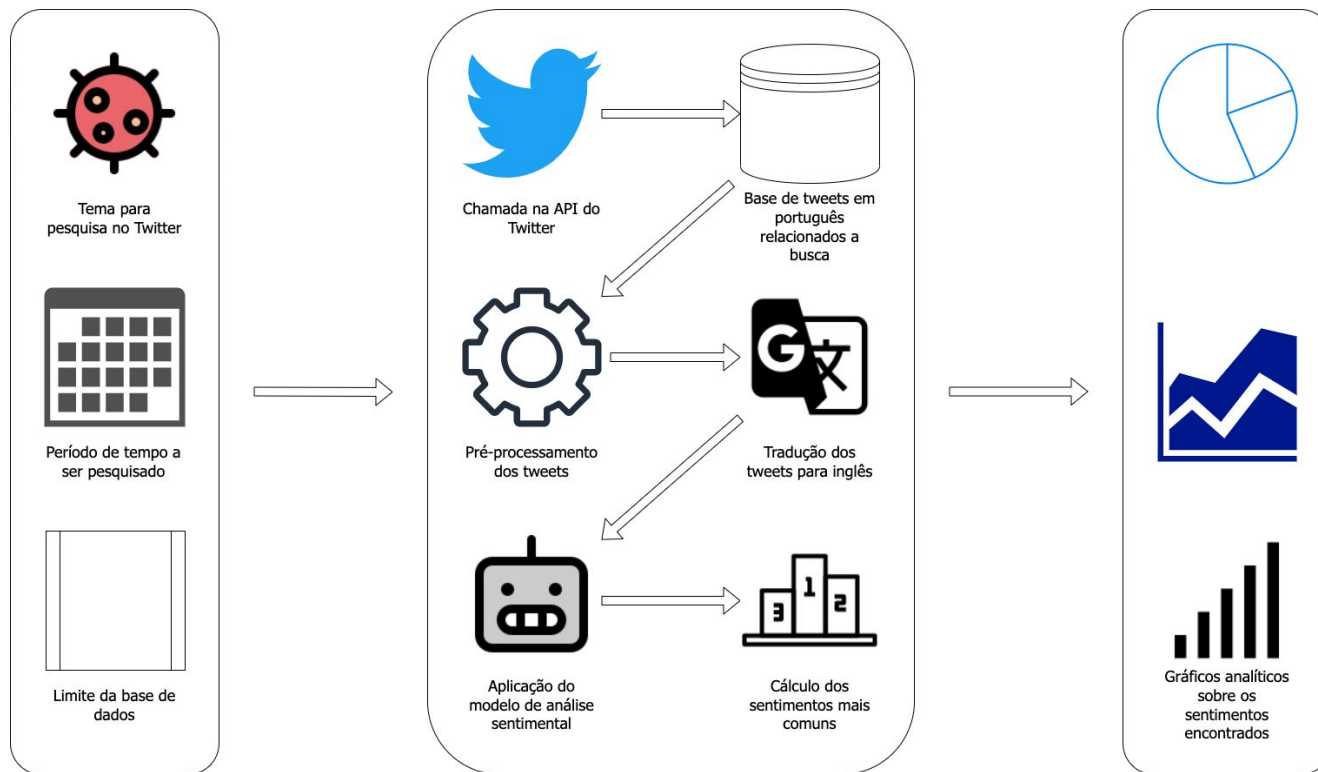
Agenda



- 01 Introdução
- 02 Trabalhos relacionados
- 03 Desenvolvimento**
- 04 Resultados
- 05 Conclusão
- 06 Referências

- Coleta de tweets em língua portuguesa
- Pré-processamento da base de textos
- Tradução dos textos para língua inglesa
- Aplicação do modelo de análise sentimental
- Criação de base de dados para geração de gráficos

Desenvolvimento: Arquitetura



Desenvolvimento: Coleta

```
import snsrape.modules.twitter as sntwitter

query = "covid OR pandemia lang:pt until:2023-01-01 since:2020-01-01"
limit = 100000

dates = []
locations = []
tweets_pt = []

for tweet in sntwitter.TwitterSearchScraper(query).get_items():
    if len(tweets_pt) >= limit:
        break
    location = tweet.user.location
    if location.find('Brasil') != -1:
        date = str(tweet.date).split('+')[0]
        dates.append(date)
        locations.append(location)
        tweets_pt.append(tweet.rawContent)
```

Desenvolvimento: Pré-processamento

```
tweets_proc = []
for tweet in tweets_pt:
    tweet_words = []
    for word in tweet.split():
        if word.startswith('@') and len(word) > 1:
            word = '@user'
        elif word.startswith('http'):
            word = 'http'
        tweet_words.append(word)
    tweets_proc.append(" ".join(tweet_words))
```

Desenvolvimento: Tradução



```
from deep_translator import GoogleTranslator  
  
tweets_en = GoogleTranslator(source='pt', target='en').translate_batch(tweets_proc)
```

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification
from scipy.special import softmax

roberta = 'cardiffnlp/twitter-roberta-base-sentiment-latest'

tokenizer = AutoTokenizer.from_pretrained(roberta)
model = AutoModelForSequenceClassification.from_pretrained(roberta)

scores = []
for tweet in tweets_en:
    encoded_tweet = tokenizer(tweet, return_tensors='pt')
    output = model(**encoded_tweet)
    output_score = output[0][0].detach().numpy()
    scores.append(softmax(output_score))
```

Desenvolvimento: Base de dados

```
import pandas as pd

labels = ['Negative', 'Neutral', 'Positive']

negative = [score[0] for score in scores]
neutral = [score[1] for score in scores]
positive = [score[2] for score in scores]
label = [labels[score.argmax()]] for score in scores]

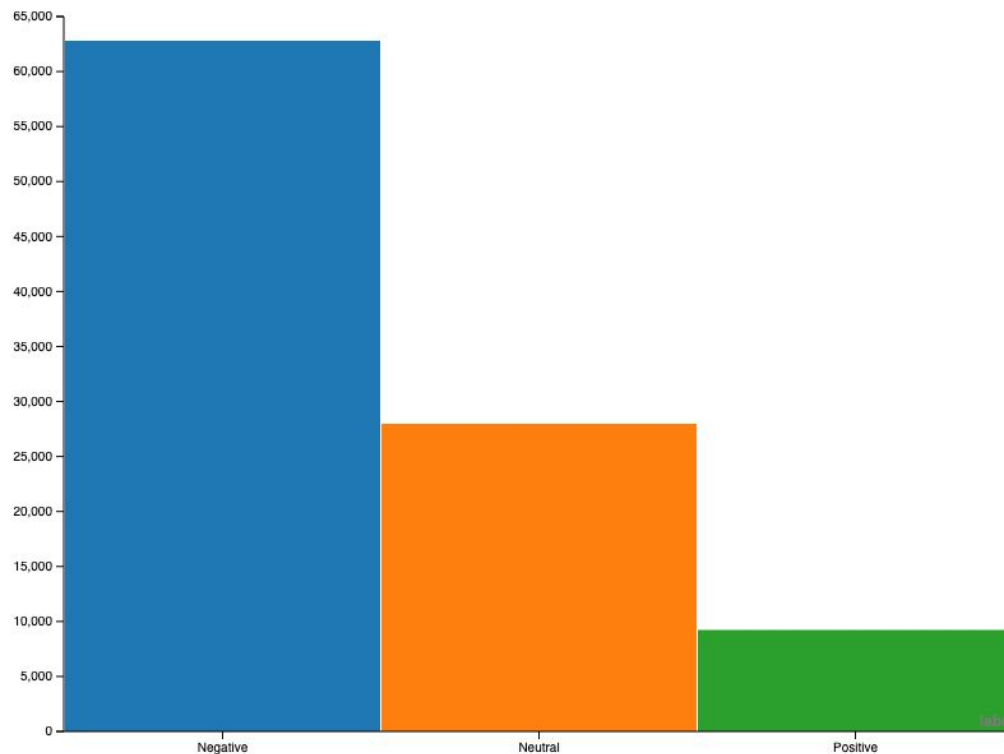
df = pd.DataFrame({
    'date': dates,
    'location': locations,
    'tweets_pt': tweets_proc,
    'tweets_en': tweets_en,
    'negative': negative,
    'neutral': neutral,
    'positive': positive,
    'label': label
})

df.to_csv('database.csv')
```

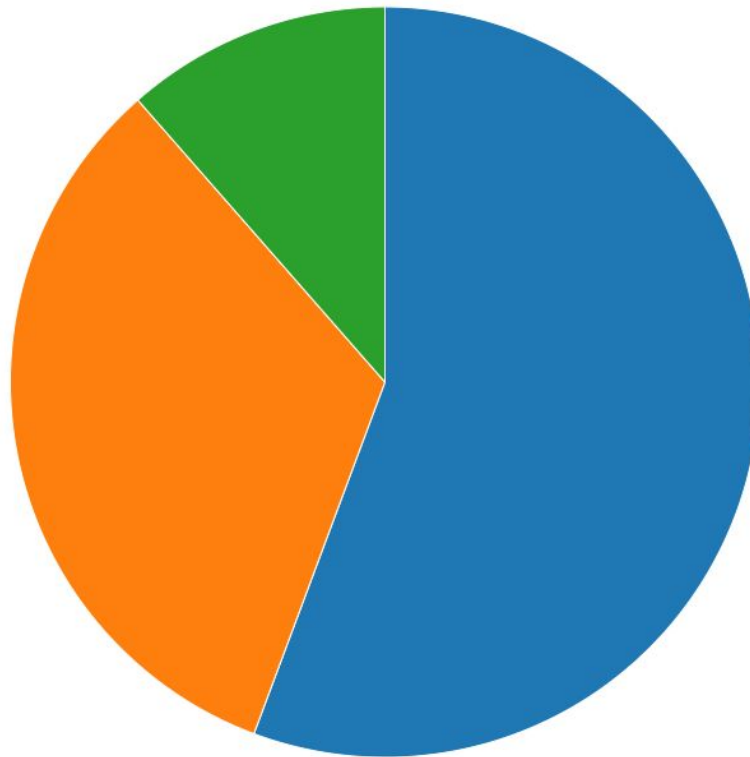

Agenda

- 01 Introdução
- 02 Trabalhos relacionados
- 03 Desenvolvimento
- 04 Resultados**
- 05 Conclusão
- 06 Referências

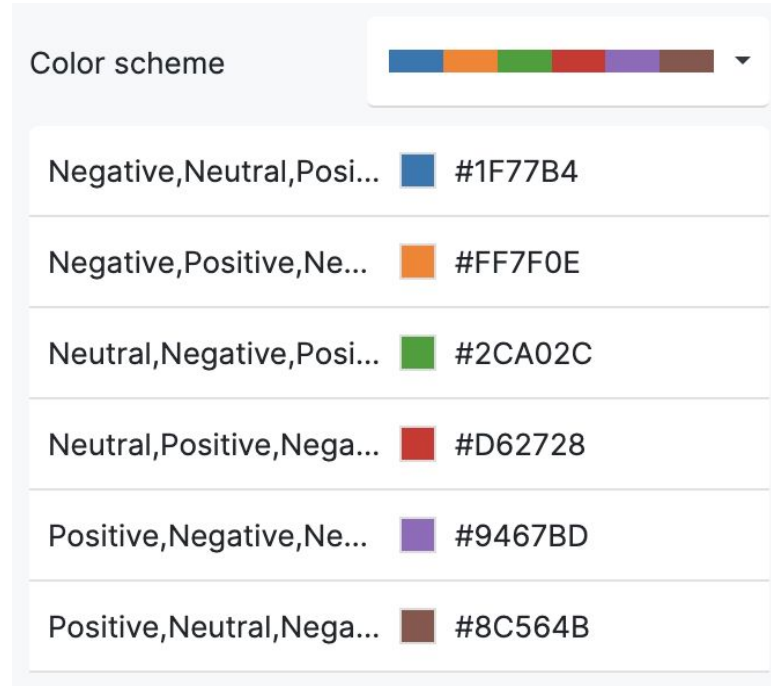
Resultados: N° de tweets por sentimiento



Resultados: Porcentagem sentimental



Resultados: Grupos de sentimientos por día



Agenda



- 01 Introdução
- 02 Trabalhos relacionados
- 03 Desenvolvimento
- 04 Resultados
- 05 Conclusão**
- 06 Referências

- Obteve-se um maior número de tweets negativos sobre os demais, seguido pelos neutros, e, por último, positivos.
- A hipótese inicial estava certa sobre uma percepção mais negativa da população sobre a covid-19.

Conclusão: Próximos passos

- Analisar sentimentalmente tweets com um modelo analítico com base na língua portuguesa.
- Analisar sentimentalmente em períodos específicos na pandemia, seja um evento marcante durante a pandemia, período com um dado que foge do padrão, etc.
- Analisar os tweets com emoções além da valência emocional, abordando raiva, surpresa, medo, etc.

Agenda



- 01 Introdução
- 02 Trabalhos relacionados
- 03 Desenvolvimento
- 04 Resultados
- 05 Conclusão
- 06 Referências**

- Mansoor et al.: Global sentiment analysis of COVID-19 tweets over time
- Melo et al.: Comparing news articles and tweets about COVID-19 in Brazil: sentiment analysis and topic modeling approach
- Sitaula et al.: Deep learning-based methods for sentiment analysis on Nepali COVID-19-related tweets
- Kausar et al.: Public sentiment analysis on Twitter data during COVID-19 outbreak
- Liu et al.: Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis

Obrigado!

Perguntas?



Artur Filgueiras Scheiba Zorron
tucazorron@gmail.com

