

análisis-de-sentimientos-a01633819

October 28, 2023

0.1 Actividad NLP - 2.0 - Aplicación de análisis de sentimientos

0.2 ##### Andrés Alejandro Guzmán González - A01633819

Importación de librerías

```
[1]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
[ ]: !pip install -q transformers
from transformers import pipeline
```

```
[3]: import pandas as pd
import matplotlib.pyplot as plt
```

Creación del dataframe con los tweets

Preda, G. (2020). Pfizer Vaccine Tweets. Recuperado de: <https://www.kaggle.com/datasets/gpreda/pfizer-vaccine-tweets/>

```
[4]: df = pd.read_csv("/content/drive/MyDrive/Archivos_Colab/vaccination_tweets.
↪ csv", encoding="latin-1")
df.head()
```

```
[4]:
```

	id	user_name	user_location \
0	1340539111971516416	Rachel Roh	La Crescenta-Montrose, CA
1	1338158543359250433	Albert Fong	San Francisco, CA
2	1337858199140118533	elið ±ð ¹ð ¢ð ¢ð	Your Bed
3	1337855739918835717	Charles Adler	Vancouver, BC - Canada
4	1337854064604966912	Citizen News Channel	NaN

	user_description	user_created \
0	Aggregator of Asian American news; scanning di...	2009-04-08 17:52:46
1	Marketing dude, tech geek, heavy metal & '80s ...	2009-09-21 15:27:30
2	heil, hydra ð â º	2020-06-25 23:30:28
3	Hosting "CharlesAdlerTonight" Global News Radi...	2008-09-10 11:28:53
4	Citizen News Channel bringing you an alternati...	2020-04-23 17:58:42

	user_followers	user_friends	user_favourites	user_verified	\
0	405	1692	3247	False	
1	834	666	178	False	
2	10	88	155	False	
3	49165	3933	21853	True	
4	152	580	1473	False	

	date	text	\
0	2020-12-20 06:06:44	Same folks said daikon paste could treat a cyt...	
1	2020-12-13 16:27:13	While the world has been on the wrong side of ...	
2	2020-12-12 20:33:45	#coronavirus #SputnikV #AstraZeneca #PfizerBio...	
3	2020-12-12 20:23:59	Facts are immutable, Senator, even when you're...	
4	2020-12-12 20:17:19	Explain to me again why we need a vaccine @Bor...	

	hashtags	source	\
0	['PfizerBioNTech']	Twitter for Android	
1	NaN	Twitter Web App	
2	['coronavirus', 'SputnikV', 'AstraZeneca', 'Pf...	Twitter for Android	
3	NaN	Twitter Web App	
4	['whereareallthesickpeople', 'PfizerBioNTech']	Twitter for iPhone	

	retweets	favorites	is_retweet
0	0	0	False
1	1	1	False
2	0	0	False
3	446	2129	False
4	0	0	False

```
[5]: df.shape
```

```
[5]: (11020, 16)
```

Extraemos solo los tweets

```
[23]: tweets = df['text'].tolist()
      tweets[:30]
```

```
[23]: ['Same folks said daikon paste could treat a cytokine storm #PfizerBioNTech
https://t.co/xeHhIMg1kF',
      "While the world has been on the wrong side of history this year, hopefully,
the biggest vaccination effort we've evâ\x80| https://t.co/dlCHrZjkhm",
      '#coronavirus #SputnikV #AstraZeneca #PfizerBioNTech #Moderna #Covid_19 Russian
vaccine is created to last 2-4 yearsâ\x80| https://t.co/ieYlCKBr8P',
      "Facts are immutable, Senator, even when you're not ethically sturdy enough to
acknowledge them. (1) You were born iâ\x80| https://t.co/jqgV18kch4",
      'Explain to me again why we need a vaccine @BorisJohnson @MattHancock']
```

#whereareallthesickpeople #PfizerBioNTechâ\x80| <https://t.co/KxbSRoBEHq>',
'Does anyone have any useful advice/guidance for whether the COVID vaccine is safe whilst breastfeeding?â\x80| <https://t.co/EifsyQoeKN>',
'it is a bit sad to claim the fame for success of #vaccination on patriotic competition between USA, Canada, UK andâ\x80| <https://t.co/IfMrAyGyTP>',
'There have not been many bright days in 2020 but here are some of the best \n1. #BidenHarris winning #Election2020â\x80| <https://t.co/77u4f8XXfx>',
'Covid vaccine; You getting it?\n\n #CovidVaccine #covid19 #PfizerBioNTech #Moderna',
'#CovidVaccine \n\nStates will start getting #COVID19Vaccine Monday, #US says \n#pakustv #NYC #Healthcare #GlobalGoalsâ\x80| <https://t.co/MksOvBvs5w>',
'while deaths are closing in on the 300,000 mark... millions of people wait #PfizerBioNTech #Vaccine !\n-\n\nThe first Uâ\x80| <https://t.co/uQ3A2f7SVP>',
'@cnbrk #COVID19 #CovidVaccine #vaccine #Corona #PfizerBioNTech #bbcnews #NYTimes #BBCNews Best wishes to the USâ\x80| <https://t.co/9xWmfU3LZj>',
'The agency also released new information for health care providers and for patients as the US shipped millions of dâ\x80| <https://t.co/sG6BtD7jD9>',
'For all the women and healthcare providers who have been asking about the safety of the #PfizerBioNTechâ\x80| <https://t.co/ow0Pglkwte>',
'"Expect 145 sites across all the states to receive vaccine on Monday, another 425 sites on Tuesday," said the officâ\x80| <https://t.co/HiUVYJzOBY>',
'Trump announces #vaccine rollout 'in less than 24 hours'\n\nThe first Americans will be vaccinated againstâ\x80| <https://t.co/2FzQSMnhoY>',
'UPDATED: #YellowFever & #COVID19 #ImmunityPassports - Part Two\n\n#SARSCoV2 #PfizerBioNTech #Britain #Decemberâ\x80| <https://t.co/qKT7Rst9aW>',
'Coronavirus: Iran reports 8,201 new cases, 221 deaths in the last 24 hours #Iran #coronavirus #PfizerBioNTechâ\x80| <https://t.co/mwDNAdmb7F>',
'.@Pfizer will rake in billions from its expensive #CovidVaccine but @AlbertBourla refuses to cut the price for loweâ\x80| <https://t.co/eC94w3TU10>',
'The trump administration failed to deliver on vaccine promises, *shocker* #COVIDIOTS #coronavirus #CovidVaccineâ\x80| <https://t.co/hew6eHTUrD>',
'How much did the #fda get paid to approve this all of a sudden now? money makes the world go around. #vaccineâ\x80| <https://t.co/GT9qVVNbKj>',
'Anyone wondering why day after #PfizerBioNTech approval in the UK people were getting vaccinated but all we are tolâ\x80| <https://t.co/tPSyL9CUYE>',
'Trump announces #vaccine rollout 'in less than 24 hours'\n\nThe first Americans will be vaccinated againstâ\x80| <https://t.co/P9esXr3zpS>',
'The US Food and Drug Administration (FDA) has granted emergency use authorization to Pfizer-BioNTech's mRNAâ\x80| <https://t.co/dPLokxGtBa>',
'Presenting you the top Medical News/Advancement of the past week you don't want to miss!\n\n#docnosofficial #covid19â\x80| <https://t.co/qOd5TPLYvA>',
'No.1 of 2 done.\n\n#ThankYouNHS @NHSuk @MHRAgovuk and #PfizerBioNTech @pfizer for making the #CovidVaccine dream possâ\x80| <https://t.co/3KvJ7UP432>',
'Wear a mask, wash your hands, and remain socially distant when possible. #stayhome #StayAtHome #StayAtHomeSaveLivesâ\x80| <https://t.co/Jrlrg6bmOw>',
'â\x81|@AvgerinosMoscowâ\x81© #PfizerBioNTech #FDA approval Apprentice

```
Style!\nâ\x81!\@realDonaldTrumpâ\x81©: â\x80\x9cApprove the #vaccine or youâ\x80\x9d
https://t.co/f64c0P5DV6',
  'Interesting and very detailed article showing up how a well tested
#supplychain, with #sharedvisibility will help dâ\x80\x9d https://t.co/yt3vn67mVg',
  "ð\x9f\x92\x89 #Vaccine #TrumpVaccine #AstraZeneca #PfizerBioNTech
#Pfizervaccine #PfizerCovidVaccine #Pfizer's COVID-19 belongâ\x80\x9d
https://t.co/ldxQhN6uix"]
```

Para eficientar el modelo reducimos el número de tweets y se entrena el modelo.

```
[7]: sentiment_analysis = pipeline("sentiment-analysis")
      modelo = sentiment_analysis(tweets[:1500])
```

No model was supplied, defaulted to distilbert-base-uncased-finetuned-sst-2-english and revision af0f99b (<https://huggingface.co/distilbert-base-uncased-finetuned-sst-2-english>).

Using a pipeline without specifying a model name and revision in production is not recommended.

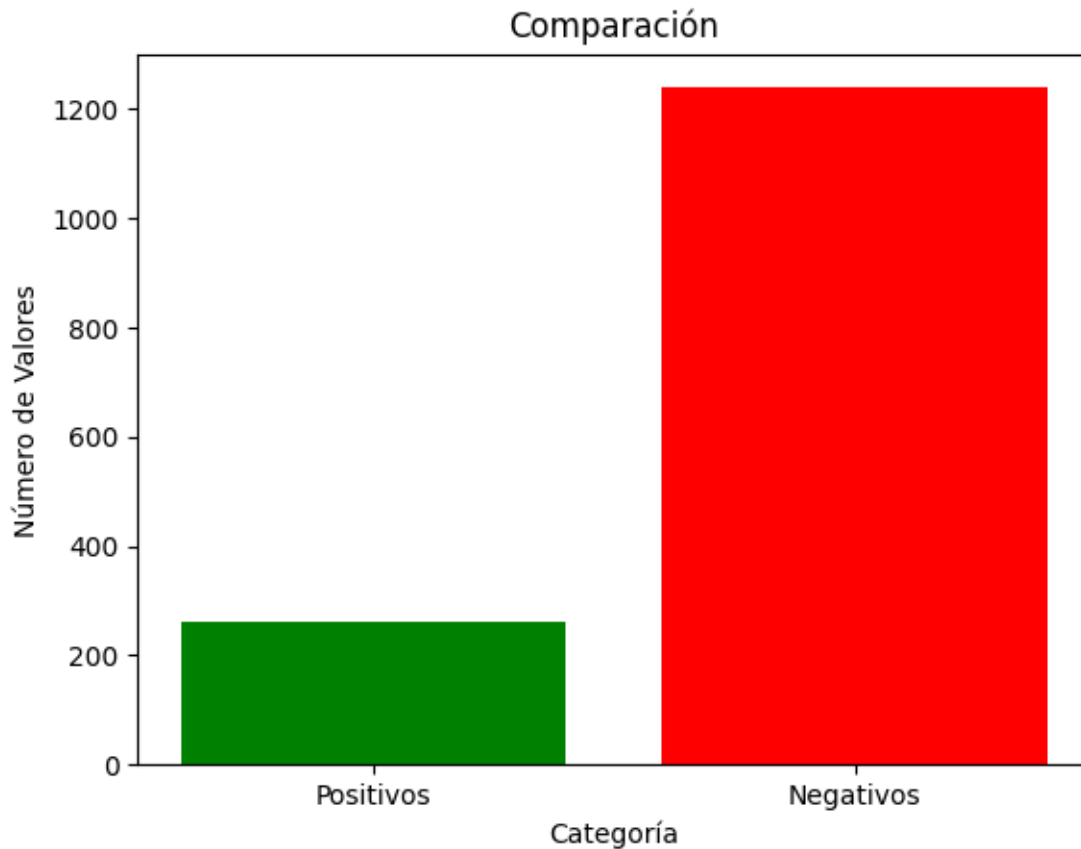
Contamos los twtets categorizados como positivos

```
[8]: positivos = 0
      negativos = 0
      for diccionario in modelo:
          valor = diccionario.get("label")
          if valor == "POSITIVE":
              positivos += 1
          else:
              negativos += 1
```

Graficación de resultados

```
[9]: etiquetas = ['Positivos', 'Negativos']
      valores = [positivos, negativos]
      plt.title("Comparación")

      plt.bar(etiquetas, valores, color=['green', 'red'])
      plt.xlabel('Categoría')
      plt.ylabel('Número de Valores')
      plt.show()
```



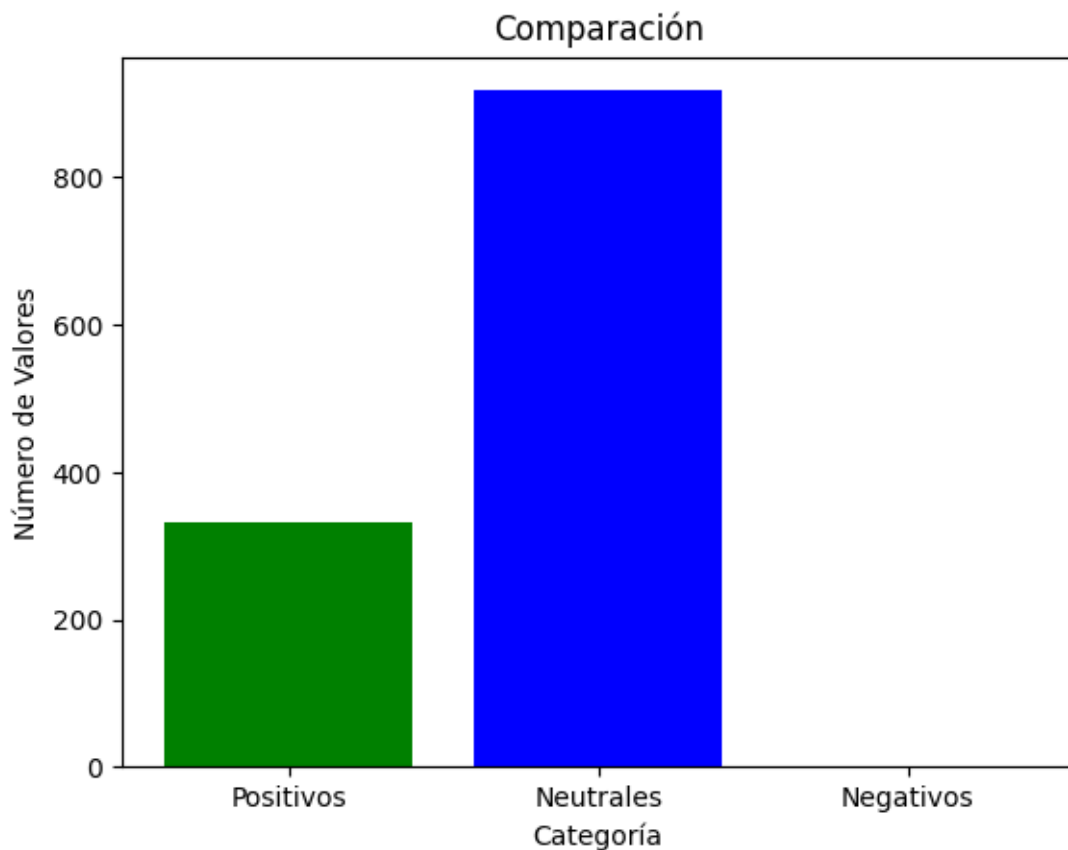
```
[10]: sentiment_analysis = pipeline(model="cardiffnlp/twitter-roberta-base-sentiment")
      modelo = sentiment_analysis(tweets[:1500])
```

```
Downloading (...)lve/main/config.json: 0%|          | 0.00/747 [00:00<?, ?B/s]
Downloading pytorch_model.bin: 0%|          | 0.00/499M [00:00<?, ?B/s]
Downloading (...)olve/main/vocab.json: 0%|          | 0.00/899k [00:00<?, ?B/s]
Downloading (...)olve/main/merges.txt: 0%|          | 0.00/456k [00:00<?, ?B/s]
Downloading (...)cial_tokens_map.json: 0%|          | 0.00/150 [00:00<?, ?B/s]
```

```
[11]: positivos = 0
      negativos = 0
      neutrales = 0
      for diccionario in modelo:
          valor = diccionario.get("label")
          if valor == "LABEL_2":
              positivos += 1
          elif valor == "LABEL_1":
              neutrales += 1
```

```
else:  
    negativos = 1
```

```
[12]: etiquetas = ['Positivos', 'Neutrales', 'Negativos']  
valores = [positivos,neutrales,negativos]  
plt.title("Comparación")  
  
plt.bar(etiquetas, valores, color=['green','blue', 'red'])  
plt.xlabel('Categoría')  
plt.ylabel('Número de Valores')  
plt.show()
```



Unimos el corpus en un único string

```
[21]: corpus_completo = '\n'.join(tweets)
```

Proceso para contar las oraciones totales de los tweets

```
[20]: cantidad_separaciones = corpus_completo.count('.')
```

```
print(f'\nCantidad de separaciones: {cantidad_separaciones}')
```

Cantidad de separaciones: 17543

Considerando que los tweets analizados fueron 1500 podemos concluir que estos son considerablemente algo.