

Solución problema Analisis de sentimiento de un texto

Dorian Felipe Marín Quintero
Carlos Miguel Rodríguez Botero

12 de diciembre 2021 Pereira, Risaralda

Análisis de sentimientos de un texto

El análisis de sentimiento es el proceso de determinar el tono emocional detrás de una serie de palabras. La herramienta de análisis de sentimiento es una técnica automatizada para extraer información significativa de los clientes, relacionada con sus actitudes, emociones y opiniones.

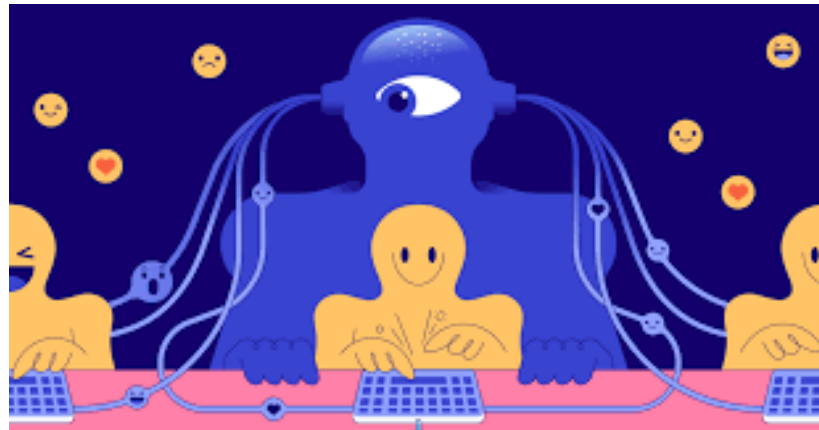
El análisis de sentimiento o sentiment analysis consiste en evaluar las emociones, actitudes y opiniones. Las organizaciones utilizan este método para obtener información que les permita comprender la forma en la que los clientes reaccionan respecto a un producto o servicio específico.



Análisis de sentimientos de un texto

La herramienta de análisis de sentimiento utiliza tecnologías avanzadas de inteligencia artificial, como procesamiento del lenguaje natural, análisis de texto y ciencia de datos para identificar, extraer y estudiar información subjetiva. En términos más simples, clasifica un texto como positivo, negativo o neutral.

Las métricas tradicionales, como el número de vistas, clics, me gusta, compartir, comentarios, etc. se centran en la cantidad. El análisis de sentimiento va más allá de los números y se centra en la calidad de las interacciones entre el público y la organización.



Análisis de sentimientos de un texto

El análisis de sentimiento del cliente puede ayudar a responder preguntas como:

- ¿Qué les gusta a nuestros clientes sobre nuestros productos y servicios?
- ¿Qué no les gusta a nuestros clientes de nuestros productos y servicios?
- ¿Recibimos demasiadas respuestas negativas recientemente?
- ¿Ha aumentado gradualmente el número de respuestas negativas?
- ¿Qué producto de marca tiene el mayor número de respuestas positivas?
- ¿El número de respuestas positivas, neutrales y negativas se ha mantenido constante en comparación con el último trimestre?
- ¿Hay un cambio en el grado de respuestas positivas o negativas?

Análisis de sentimientos de un texto

Usos de la herramienta de análisis de sentimientos



Problema número 4

Sistema que es capaz de obtener texto de una fuente externa. El objetivo es reconocer patrones de lenguaje, de modo que se pueda establecer la naturaleza del texto. Reconocer:

- A) El tema
- B) Personajes
- C) Si el texto es positivo, negativo o neutro

Librerías

- **Librería NLTK:** El Natural Language Toolkit (NLTK) es una plataforma usada para construir programas para análisis de texto. La plataforma fue liberada originalmente para análisis de texto. La plataforma fue liberada originalmente por Steven Bird y Edward Loper en conjunto con un curso de lingüística computacional en la Universidad de Pennsylvania en 2001. Hay un libro de acompañamiento para la plataforma llamado Procesamiento de Lenguaje Natural con Python.
- **Librería textblob:** textblob es una librería de procesamiento del texto para Python que permite realizar tareas de Procesamiento del Lenguaje Natural como análisis morfológico, extracción de entidades, análisis de opinión, traducción automática, etc.

Librerías

- A continuación se hace uso de la **librería transformers** la cual es una librería de python (comunmente usada en Machine learning). Esta proporciona modelos preentrenados particularmente en Texto, audio y video. Esta vez vamos a hacer uso de ella con un enfoque en Texto. :
- En texto se usa principalmente para tareas como clasificación de los mismos, extracción de información, respuesta a preguntas, resumen, traducción, generación de textos, en más de 100 idiomas.
- Con la parte de la respuesta a preguntas trabajaremos para responder a los puntos 2 y 3 del problema elegido (anteriormente mencionado). Es decir, vamos a encontrar los personajes mencionados en un texto y vamos a clasificar el tipo de texto es el que se esta trabajando.

Código

▶ `pip install nltk`



Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (3.2.5)

Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from nltk) (1.15.0)



```
import nltk
nltk.download('vader_lexicon') #De la libreria nltk se descargarán 2 componentes
nltk.download('punkt')        #indispensables para el analisis de sentimientos
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
[nltk_data] Package vader_lexicon is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
True
```

Código

```
[ ] tokenizer = nltk.data.load('tokenizers/punkt/english.pickle') #Trocear un texto con varias ideas con un proceso de tokenización
sentences = tokenizer.tokenize("I am sure that is the reason why education is so important. When you are truly interested in something, you never stop learning. Later, do not stop to sharing knowledge with others.")
#En Sentences se dividen las 3 frases del texto.
```

```
from nltk.sentiment.vader import SentimentIntensityAnalyzer #Con estos componentes se analizara cada una de las 3 frases anteriormente
from nltk import sentiment #obtenidas
from nltk import word_tokenize
```

```
/usr/local/lib/python3.7/dist-packages/nltk/twitter/__init__.py:20: UserWarning: The twython library has not been installed. Some functionality from the twitter package will not be available.
warnings.warn("The twython library has not been installed. ")
```

```
[ ] analizador = SentimentIntensityAnalyzer() #Creando el analizador de sentimientos

#Se analiza cada una de las frases guardadas en "sentences"

for sentence in sentences:
    print(sentence)
    scores = analizador.polarity_scores(sentence)
    for key in scores:
        print(key, ': ', scores[key])
    print()

#Criterios para la puntuación (score)

#Negativa : Valor entre 0 y 1 -> 0 es lo más negativo
#Positiva : Valor entre 0 y 1 -> 1 es lo más positivo
#Neutra : Valor entre 0 y 1 -> 0.5 es el puntaje más neutral
#Compound: Valor entre -1 y 1 -> -1 negativo, 0 neutral, 1 positivo
```



I am sure that is the reason why education is so important.

neg : 0.0

neu : 0.659

pos : 0.341

compound : 0.567

When you are truly interested in something, you never stop learning.

neg : 0.0

neu : 0.517

pos : 0.483

compound : 0.7571

Later, do not stop to sharing knowledge with others.

neg : 0.208

neu : 0.624

pos : 0.168

compound : -0.1139

Código



Código

```
[ ] from textblob import TextBlob #Hacemos uso de Textblob, este nos devuelve 2 propiedades

[ ] text = TextBlob("Estoy haciendo una prueba de TextBlob y parece que funciona bien") #Se escribe la frase a analizar
    text2 = TextBlob("Estoy siendo totalmente objetivo en el analisis")

    text_EnglishVersion = text.translate(to="en") #Se traduce la frase de español a ingles
    text2_EnglishVersion = text2.translate(to="en")

    print("La frase #1 en ingles es: ", text_EnglishVersion) #Muestra un mensaje de la traducción que se hace
    print("La frase #2 en ingles es: ", text2_EnglishVersion, "\n")

    sentimiento = text_EnglishVersion.sentiment #Hace el analisis de frase #1 y muestra en pantalla
    print("Resultados frase #1",sentimiento)

    sentimiento2 = text2_EnglishVersion.sentiment #Hace el analisis de la frase #2 y muestra en pantalla
    print("Resultados frase #2",sentimiento2)
```

La frase #1 en ingles es: I'm doing a TextBlob test and it seems to work fine
La frase #2 en ingles es: I am being totally objective in the analysis

Resultados frase #1 Sentiment(polarity=0.4166666666666667, subjectivity=0.5)
Resultados frase #2 Sentiment(polarity=0.0, subjectivity=0.1)

Código

```
[ ] from transformers import pipeline
```

```
[ ] modelo = pipeline(task="question-answering");
```

```
texto = r""" The film is based on the science fiction genre book and narrates the journey of the main protagonist, Frodo Baggins,
hobbit of the Shire, to destroy the One Ring and the subsequent war that will provoke the enemy to recover it, as it is the main
source of power of its creator, the Dark Lord Sauron."""
```

```
print(modelo(question="Who are characters in the book?", context=texto))
print(modelo(question="What is the book genre?", context=texto))
print(modelo(question="Who is the dark lord?", context=texto))
```

No model was supplied, defaulted to distilbert-base-cased-distilled-squad (<https://huggingface.co/distilbert-base-cased-distilled-squad>)

```
{'score': 0.9126824736595154, 'start': 103, 'end': 116, 'answer': 'Frodo Baggins'}
{'score': 0.9738324880599976, 'start': 26, 'end': 41, 'answer': 'science fiction'}
{'score': 0.9827685356140137, 'start': 299, 'end': 305, 'answer': 'Sauron'}
```

Bibliografía

- <https://code.tutsplus.com/es/tutorials/introduciendo-the-natural-language-toolkit-nltk--cms-28620>
- <https://huggingface.co/docs/transformers/index>
- <https://colab.research.google.com/github/vitojph/kschool-nlp-18/blob/master/notebooks/textblob.ipynb#scrollTo=PTEKFmuU8E-H>

MUCHAS GRACIAS