

ANÁLISIS DE SENTIMIENTO

ANALÍTICA WEB

MÁSTER UNIVERSITARIO
INTELIGENCIA COMPUTACIONAL E INTERNET DE LAS COSAS

Amelia Zafra Gómez



- 1. Introducción**
2. Aplicaciones y medidas de evaluación
3. Métodos basados en léxico
4. Entorno y librerías de Python
5. Google Cloud Natural Language API

1. Introducción

- * **El Análisis de Sentimientos**
 - * Combina estadística, el procesamiento del lenguaje natural y técnicas de Machine Learning.
 - * Asigna significado emocional a las comunicaciones.



1. Introducción

- * **Funcionalidad**

- * Permite medir la opinión de los clientes a escala de forma automática.



1. Introducción

* Principales desafíos

**«Entre lo que pienso,
lo que quiero decir,
lo que creo decir,
lo que digo,
lo que quieres oír,
lo que oyes,
lo que crees entender,
lo que entiendes,
existen nueve posibilidades de no
entenderse».**

1. Introducción

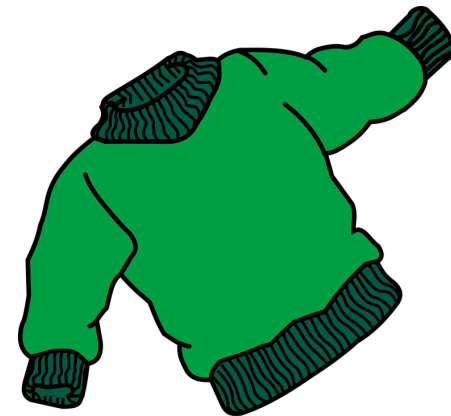
- * Principales desafíos

- * Subjetividad

Ese jersey es espectacular 🤔

Odio ese jersey 😡

El jersey es verde 😐



1. Introducción

- * Principales desafíos

- * Manejo de la negación

Absolutamente nada 🤔

Todo de él 🤔

¿Qué te gustó del congreso?



¿Qué no te gustó del congreso?

1. Introducción

- * **Principales desafíos**

- * Ironía y sarcasmo

Siento tanta felicidad al
saber que tengo que esperar
8 horas más por el retraso
del vuelo

Las odio!! Pero me encantan
como lucen



1. Introducción

- * **Principales desafíos**

- * Definiendo lo neutral

Textos objetivos: los llamados textos objetivos no contienen sentimientos explícitos, por lo que se debe incluir esos textos en la categoría neutral.

Información irrelevante: La información procesada previamente y que no se considera relevante en el estudio, se puede etiquetar como neutral.

Textos que contengan deseos: Algunos deseos como *me gustaría que el producto tuviera más chocolate* son generalmente neutrales. Sin embargo, aquellos textos que incluyen comparaciones, como *me gustaría que el producto fuera mejor*, son bastante difíciles de categorizar.

1. Introducción

- * **Principales desafíos**

- * Matices, tono, variaciones culturales, jergas, ...

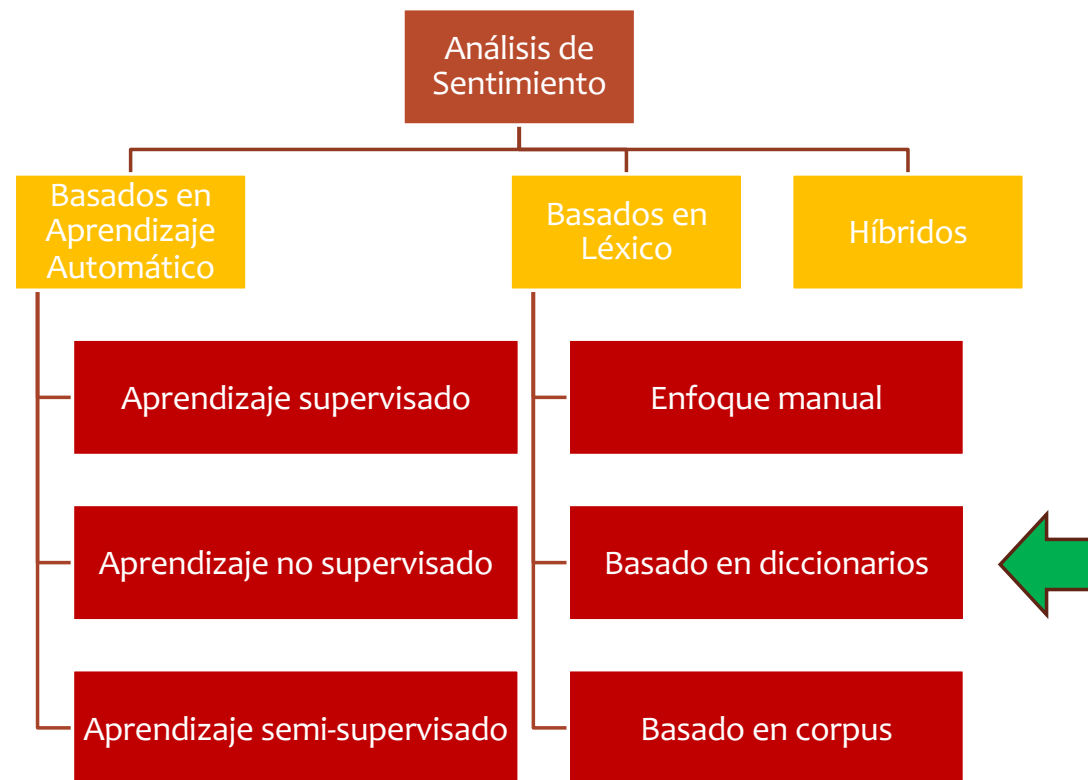
La segunda vez que mi vuelo se retrasa. ¡Genial!

Los helados me vuelven loco.

Hay que estar loco para comerse uno de esos helados.

1. Introducción

* Enfoques



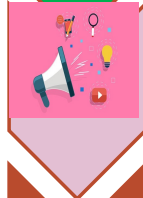
Índice

1. Introducción
- 2. Aplicaciones y medidas de evaluación**
3. Métodos basados en léxico
4. Entorno y librerías de Python
5. Google Cloud Natural Language API

2. Aplicaciones y medidas de evaluación



- Inteligencia de negocio



- Sistemas de Recomendación



- Inteligencia de gobierno



- Salud y dominio médico

2. Aplicaciones y medidas de evaluación

* Matriz de confusión

		Predicción	
		Positiva	Negativa
Real	Positiva	Verdaderos positivos (TP)	Falsos negativos (FN)
	Negativa	Falsos positivos (FP)	Verdaderos negativos (TN)

2. Aplicaciones y medidas de evaluación

* Matriz de confusión

		Predicción		
		Positivo	Neutro	Negativo
Real	Positivo	Textos con sentimiento positivos clasificados correctamente	Textos con sentimiento positivo clasificados como neutros	Textos con sentimiento positivo clasificados como negativos
	Neutro	Textos con sentimiento neutros clasificados como positivos	Textos con sentimiento neutros clasificados correctamente	Textos con sentimiento neutros clasificados como negativos
	Negativo	Textos con sentimiento negativo clasificados como positivos	Textos con sentimiento negativo clasificados como neutros	Textos con sentimiento negativo clasificados correctamente

2. Aplicaciones y medidas de evaluación

* Matriz de confusión

		Predicción		
		Positivo	Neutro	Negativo
Real	Positivo	7	1	3
	Neutro	8	2	2
	Negativo	9	3	1

sentimiento positivo - positivo

- $TP = 7$
- $TN (2+3+2+1) = 8$
- $FP = (8+9) = 17$
- $FN = (1+3) = 4$

sentimiento neutro - positivo

- $TP = 2$
- $TN (7+1+9+3) = 20$
- $FP = (1+3) = 4$
- $FN = (8+2) = 10$

sentimiento negativo - positivo

- $TP = 1$
- $TN (7+2+8+1) = 18$
- $FP = (2+3) = 5$
- $FN = (9+3) = 12$

2. Aplicaciones y medidas de evaluación

* Precisión

$$P = \frac{TP}{FP + TP}$$

Precisión (clase sentimiento positivo) = $(7)/(17+7) = 0.292$

Precisión (clase sentimiento neutro) = $(2)/(2+4) = 0.333$

Precisión (clase sentimiento negativo) = $(1)/(1+5) = 0.167$

sentimiento positivo - positivo

- TP = 7
- TN (2+3+2+1) = 8
- FP = (8+9) = 17
- FN = (1+3) = 4

sentimiento neutro - positivo

- TP = 2
- TN (7+1+9+3) = 20
- FP = (1+3) = 4
- FN = (8+2) = 10

sentimiento negativo - positivo

- TP = 1
- TN (7+2+8+1) = 18
- FP = (2+3) = 5
- FN = (9+3) = 12

2. Aplicaciones y medidas de evaluación

* Precisión

$$P = \frac{TP}{FP + TP}$$

Micro-average precisión

- Total TP = 7 + 2 + 1 = 10
- Total TN = 8 + 20 + 18 = 46
- Total FN = 4 + 10 + 12 = 26
- Total FP = 17 + 4 + 5 = 26

Micro precisión = $(10) / (10 + 26) = 0.278$

sentimiento positivo - positivo

- TP = 7
- TN (2+3+2+1) = 8
- FP = (8+9) = 17
- FN = (1+3) = 4

sentimiento neutro - positivo

- TP = 2
- TN (7+1+9+3) = 20
- FP = (1+3) = 4
- FN = (8+2) = 10

sentimiento negativo - positivo

- TP = 1
- TN (7+2+8+1) = 18
- FP = (2+3) = 5
- FN = (9+3) = 12

2. Aplicaciones y medidas de evaluación

* Precisión

$$P = \frac{TP}{FP + TP}$$

Macro-average precisión

Precisión (clase sentimiento positivo) =
(7)/(17+7) = 0.292

Precisión (clase sentimiento neutro) = (2)/(2+4)
= 0.333

Precisión (clase sentimiento negativo) =
(1)/(1+5) = 0.167

Macro precisión = (0.292 + 0.333 + 0.167) / 3 =
0.264

sentimiento positivo - positivo

- TP = 7
- TN (2+3+2+1) = 8
- FP = (8+9) = 17
- FN = (1+3) = 4

sentimiento neutro - positivo

- TP = 2
- TN (7+1+9+3) = 20
- FP = (1+3) = 4
- FN = (8+2) = 10

sentimiento negativo - positivo

- TP = 1
- TN (7+2+8+1) = 18
- FP = (2+3) = 5
- FN = (9+3) = 12

Índice

1. Introducción
2. Aplicaciones y medidas de evaluación
- 3. Métodos basados en léxico**
4. Entorno y librerías de Python
5. Google Cloud Natural Language API

3. Métodos basados en léxico

- * **Taxonomía**

- * **Enfoque manual**

SentiWordNet

WordNet-AFFECT

Q-WordNet

- * **Enfoque basado en diccionarios**

Léxico de Bing Liu

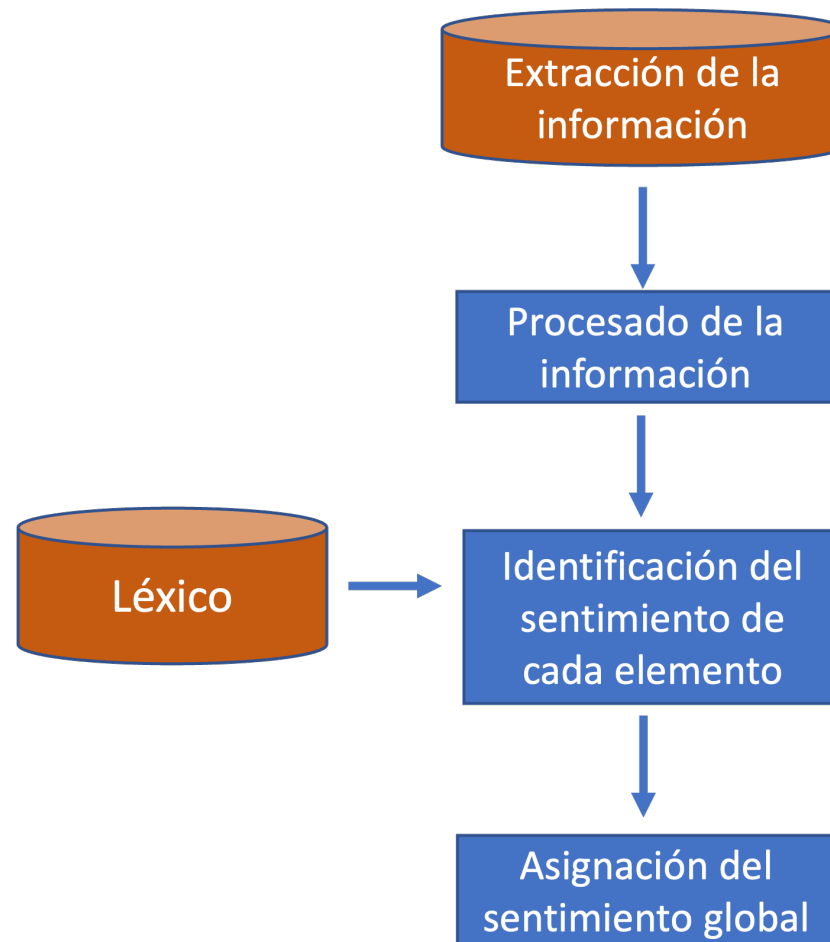
- * **Enfoque basado en corpus**

ElhPolar

ISOL

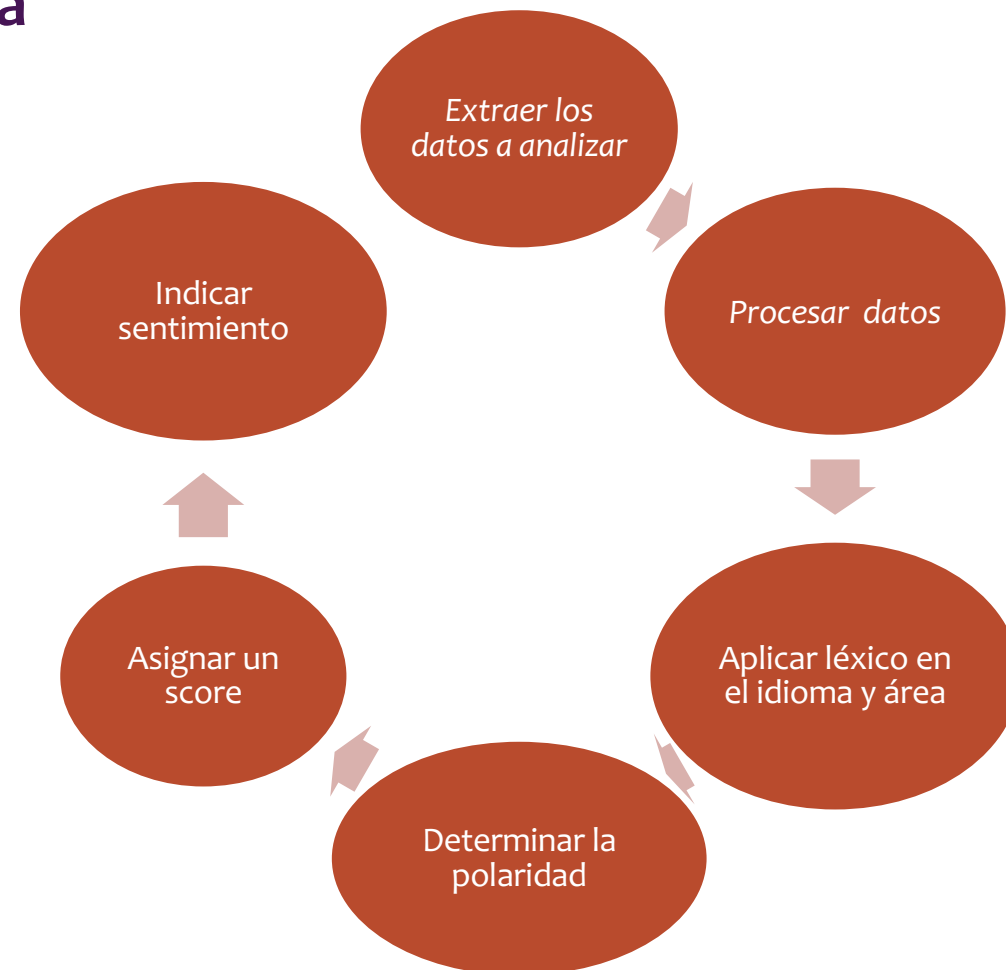
3. Métodos basados en léxico

* Estructura



3. Métodos basados en léxico

* Metodología



3. Métodos basados en léxico

- * Metodología

- * Extraer la información



3. Métodos basados en léxico

- * **Metodología**

- * **Procesar la información**

Tokenización

Lematización

Filtrado de
palabras

Stemming

3. Métodos basados en léxico

* Metodología

* Procesar la información

Tokenización

Dividir el texto en palabras

Se debe eliminar: símbolos de puntuación, errores ortográficos, mayúsculas o minúsculas...

	este	es	un	ejemplo	de	dividir	un	texto
Texto1	1	1	1	1	0	0	0	1
Texto2	1	1	1	1	1	1	1	1


Métodos basados en léxico

* Metodología

* Procesar la información

Filtrado de palabras

Eliminar palabras vacías (sin significado relevante): artículos, conjunciones, preposiciones y pronombres



	este	es	un	ejemplo	de	dividir	un	texto
Texto1	1	1	1	1	0	0	0	1
Texto2	1	1	1	1	1	1	1	1

Métodos basados en léxico

* Metodología

* Procesar la información

Filtrado de
palabras

Otras técnicas de filtrado más elaboradas contemplan:

- Eliminar palabras que no aportan nada en un contexto específico.
- Sustituir palabras por un texto alternativo común.

Métodos basados en léxico

- * Metodología

- * Procesar la información

Stemming

Reducir los términos a su raíz

Zapato
Zapatería
Zapatero

Métodos basados en léxico

- * Metodología

- * Procesar la información

Lematización

Reducir variantes morfológicas de las formas de una palabra a raíces comunes o lexemas:

- plural, femenino, conjugada...

La lematización utiliza diccionarios para encontrar la base de la palabra, lo que también se conoce como lema de las formas flexionadas.

Caminando → camin

Mejores → mejor



Caminando → caminar

Mejores → bueno

3. Métodos basados en léxico

* Aplicar el léxico

```
<layer level="1">
  <positive>
    <lemma pos="a" pol="0.708" std="0.149"> acertado </lemma>
    <lemma pos="a" pol="0.906" std="0.125"> admirable </lemma>
    <lemma pos="n" pol="0.45" std="0.331"> admiración </lemma>
    <lemma pos="v" pol="0.75" std="0.177"> admirar </lemma>
    <lemma pos="a" pol="0.375" std="0.0"> afectivo </lemma>
    <lemma pos="n" pol="0.321" std="0.112"> afecto </lemma>
    <lemma pos="a" pol="0.563" std="0.131"> afectuoso </lemma>
    <lemma pos="n" pol="0.5" std="0.425"> afición </lemma>
    <lemma pos="a" pol="0.813" std="0.063"> afortunado </lemma>
    <lemma pos="a" pol="0.75" std="0.181"> agradable </lemma>
    <lemma pos="a" pol="0.5" std="0.125"> agradecido </lemma>
    <lemma pos="v" pol="0.458" std="0.315"> alegrar </lemma>
    <lemma pos="a" pol="0.661" std="0.185"> alegre </lemma>
    <lemma pos="a" pol="0.594" std="0.0"> alentador </lemma>
    <lemma pos="v" pol="0.275" std="0.418"> alentar </lemma>
    <lemma pos="a" pol="0.688" std="0.165"> amable </lemma>
    <lemma pos="v" pol="1.0" std="0.0"> amar </lemma>
    <lemma pos="a" pol="0.708" std="0.125"> amistoso </lemma>
    <lemma pos="n" pol="0.278" std="0.171"> amor </lemma>
    <lemma pos="a" pol="0.494" std="0.276"> animado </lemma>
    <lemma pos="v" pol="0.289" std="0.232"> animar </lemma>
    <lemma pos="a" pol="0.313" std="0.161"> apacible </lemma>
    <lemma pos="a" pol="0.469" std="0.12"> apasionado </lemma>
    <lemma pos="n" pol="0.375" std="0.0"> apego </lemma>
```

a_ciegas	negative
a_flote	positive
a_la_deriva	negative
a_la_moda	positive
a_la_sombra	negative
a_pesar_de	negative
a_salvo	positive
abandonado	negative
abandonar	negative
abandono	negative
abaratar	negative
abatido	negative
abatirse	negative
abdicar	negative
aberracion	negative
abofetear	negative
abogar_por	positive
abominable	negative
abominacion	negative
abordar	negative
aborrecer	negative
aborrecimiento	negative
abrasador	negative
abrazar	positive
abrazarse	positive
abrazo	positive
abrumar	negative
absolver	positive
absorbente	positive

3. Métodos basados en léxico

* Aplicar el léxico

Texto	Positivas	Negativas
Imposible ponerme en contacto con vosotros. Qué horror , que inseguridad da su compañía	0	3
Creo que siendo una buena empresa la publicidad no inadecuada	1	1
Gracias, ya fue resuelto de manera oportuna y eficiente	2	0

3. Métodos basados en léxico

* Determinar la polaridad

Texto	Positivas	Negativas	Positivas - Negativas	Polaridad
Imposible ponerme en contacto con vosotros. Qué horror , que inseguridad da su compañía	0	3	-3	Negativa
Creo que siendo una buena empresa la publicidad no inadecuada	1	1	0	Neutra
Gracias, ya fue resuelto de manera oportuna y eficiente	2	0	2	Positiva

3. Métodos basados en léxico

* Asignar un score

Texto	Positivas - Negativas	<u>Retuits</u>	Favoritos	Polaridad	Score
Imposible ponerme en contacto con vosotros. Qué horror , que inseguridad da su compañía	-3	8	1	Negativa	$-3-(8/2)-1=-8$
Creo que siendo una buena empresa la publicidad no inadecuada	0	2	1	Neutra	0
Gracias, ya fue resuelto de manera oportuna y eficiente	2	6	3	Positiva	$2+(6/2)+3=8$

3. Métodos basados en léxico

* Indicar el sentimiento

$$\text{Sentimiento} = \begin{cases} \text{Si } Score \geq 6; & \text{Excepcional} \\ \text{Si } 1 < Score < 6; & \text{Muy bueno} \\ \text{Si } 0 < Score \leq 1; & \text{Bueno} \\ \text{Si } Score = 0; & \text{Neutral} \\ \text{Si } -1 \leq Score < 0; & \text{Malo} \\ \text{Si } -6 < Score < -1; & \text{Muy malo} \\ \text{Si } Score \leq -6; & \text{Horrible} \end{cases}$$

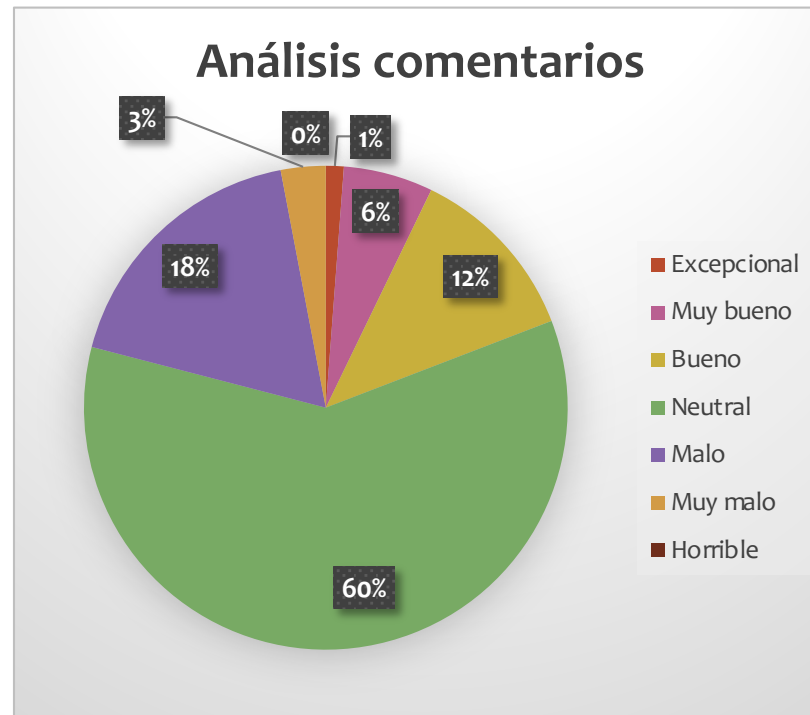
3. Métodos basados en léxico

* Indicar el sentimiento

Texto	Score	Sentimiento
Imposible ponerme en contacto con vosotros. Qué horror , que inseguridad da su compañía	-8	Horrible
Creo que siendo una buena empresa la publicidad no inadecuada	0	Neutral
Gracias, ya fue resuelto de manera oportuna y eficiente	8	Excepcional

3. Métodos basados en léxico

* Análisis



Índice

1. Introducción
2. Aplicaciones y medidas de evaluación
3. Métodos basados en léxico
- 4. Entorno y librerías de Python**
5. Google Cloud Natural Language API

4. Entornos y librerías de Python 1

- * Librerías de Python – pandas

<https://pandas.pydata.org/>

- * la **readcsv**

Salida:

>> ID	NOMBRE	EDAD
>> 0	Ana	20
>> 1	Sara	15
>> 2	Luis	18
>> 3	Pedro	24

```
#Importar la librería pandas
import pandas as pd
#Leer un archivo en csv y
#pasarlo a dataframe
df = pd.read_csv("dataset.csv")
#Imprimir el dataframe
print(df)
```

4. Entornos y librerías de Python 1

- * Librerías de Python – nltk

<https://www.nltk.org/>

- * La función **download()**:

```
#Importar el recurso  
from nltk.corpus import stopwords
```

```
#Descargar un recurso  
nltk.download('stopwords')
```


4. Entornos y librerías de Python

```
text = "Los delfines son un tipo de mamíferos cetáceos pertenecientes a la familia Delphinidae. Pueden vivir más de 30 años en cautividad y tienen una longitud de 3,5 metros aproximadamente. En la naturaleza, estos nadadores elegantes pueden alcanzar velocidades de más de 30 kilómetros por hora. Durante sus desplazamientos surgen a menudo a la superficie del mar para respirar, haciéndolo una media de dos o tres veces por minuto. Viajan en grupos sociales y se comunican entre sí por un complejo sistema de chirridos y silbidos que del mismo modo les sirve para ecolocalizar a sus presas. Pueden producir hasta 1.000 ruidos por segundo. Estos sonidos viajan bajo el agua hasta que encuentran objetos, luego regresan a sus remitentes de delfines, revelando la ubicación, tamaño y forma de su objetivo. Se trata de animales muy inteligentes que no dejan de sorprender a los científicos pues han mostrado capacidades de aprendizaje y cognición muy superiores a las de otras especies."
```

4. Entornos y librerías de Python

* Librerías de Python – scikit-learn

<https://scikit-learn.org/stable/>

* `sklearn.metrics.confusion_matrix(y_true, y_pred)`

```
#Importar las librerías
from sklearn.metrics
import confusion_matrix

#Valores reales y predichos
y_true = [2, 0, 2, 2, 0, 1]
y_pred = [0, 0, 2, 2, 0, 2]

#Calcula la matriz de confusión
confusion_matrix(y_true, y_pred)

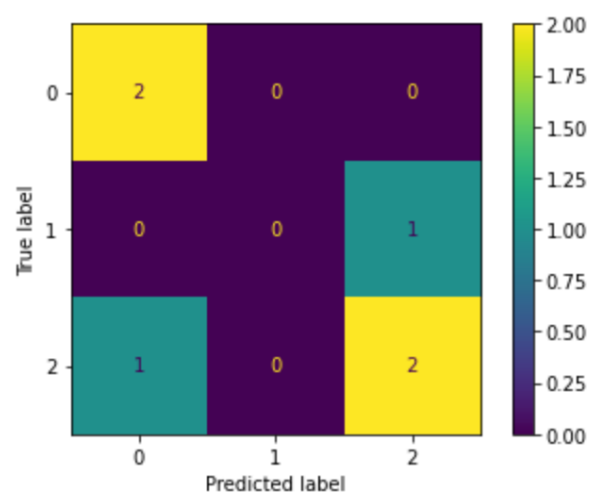
>>> array([[2, 0, 0],
          [0, 0, 1],
          [1, 0, 2]])
```

4. Entornos y librerías de Python

* Librerías de Python – scikit-learn

<https://scikit-learn.org/stable/>

* `sklearn.metrics.ConfusionMatrixDisplay.from_predictions(y_true, y_pred)`



```
#Importar las libreríasimport  
matplotlib.pyplot as plt
```

```
from sklearn.metrics import  
ConfusionMatrixDisplay
```

```
#Valores reales y predichos
```

```
y_true = [2, 0, 2, 2, 0, 1]
```

```
y_pred = [0, 0, 2, 2, 0, 2]
```

```
#Mostramos la matriz de  
confusion
```

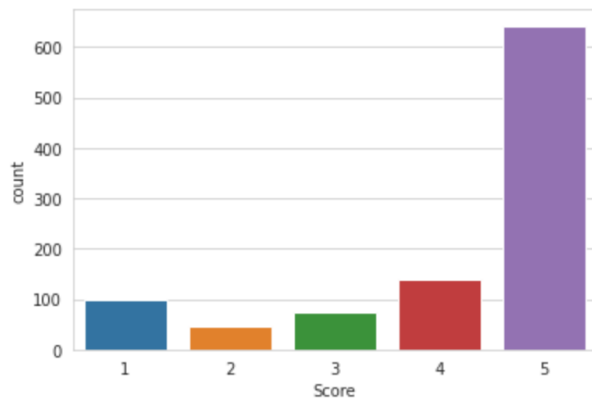
```
ConfusionMatrixDisplay.from_prediction  
s(y_true, y_pred)plt.show()
```

4. Entornos y librerías de Python

- * Librería seaborn

<https://seaborn.pydata.org/>

- * **countplot()**



```
# Importar las librerías
import pandas as pd
import seaborn as sns
#Leer datos de un csv
df_reviews = pd.read_csv('Reviews-
red.csv')
#Mostrar un gráfico de barra, de acuerdo
#al atributo score
sns.countplot(x='Score',data=df_reviews)
```

4. Entornos y librerías de Python

- * Librerías TextBlob

<https://textblob.readthedocs.io/en/dev/>

- * **sentiment**

```
from textblob import TextBlob
text = TextBlob("Textblob is amazingly simple to use.
What great fun!")
#Obtener polaridad y subjetividad
text.sentiment
```

```
>>> Sentiment (polarity=0.39166666666666666,
                subjectivity=0.4357142857142857)
```

4. Entornos y librerías de Python

- * Librería TextBlob

<https://textblob.readthedocs.io/en/dev/>

- * **sentiment.polarity:** devuelve solamente la polaridad.
- * **sentiment.subjectivity:** devuelve solamente la subjectividad.
- * **translate():** permite traducir el texto para trabajar en otros idiomas.

4. Entornos y librerías de Python

- * Librerías TextBlob

<https://textblob.readthedocs.io/en/dev/>

- * **tags:** permite acceder a la especificación del lenguaje (POS).
- * **words:** permite separar el texto en palabras.
- * **sentences:** permite separar el texto en frases.
- * **lemmatize():** realiza el proceso de lematización.

Índice

1. Introducción
2. Aplicaciones y medidas de evaluación
3. Métodos basados en léxico
4. Entorno y librerías de Python
5. **Google Cloud Natural Language API**

5. Google Cloud Natural Language API

- * <https://cloud.google.com/natural-language?hl=es>
- * Natural Language es una herramienta para **analizar el sentimiento, el lenguaje y las frases** que contiene **cualquier tipo de texto**.
 - * Obtener una lista de entidades.
 - * Obtener comentarios positivos o negativos.
- * Crear credenciales de Google Cloud Platform

5. Google Cloud Natural Language API

- * Sentimiento
 - * **Puntuación.** Nos devuelve un número con decimales entre -1 y 1, indicando -1 un sentimiento muy negativo y 1 un sentimiento muy positivo.
 - * **Magnitud.** Nos devuelve un número con decimales entre 0 y infinito. Esta variable no se normaliza, cada expresión de emoción en el texto contribuye en la magnitud.
 - * Sirve mucho para comparaciones de textos con similar longitud o evoluciones del mismo (ejemplo: cómo evoluciona la magnitud de los comentarios de tu red social).

ANÁLISIS DE REDES SOCIALES

ANALÍTICA WEB

