



CIENCIA DE DATOS II

Análisis de chats con Python

Anthony Mauricio Jiménez Navarro
Javier Antonio Hernández Navarro
Luis Fernando Amey Apuy

Agenda

Introducción	3	Planteamiento del programa	10
Nuestra motivación	4	Desarrollo del módulo	11
Nuestros objetivos	5	Resultados	14
Camino a seguir	6	Conclusiones	28
Librerías de Python	7	Recomendaciones	28

Análisis de Texto

- Proceso que permite examinar lenguaje natural.
- Permite analizar grandes cantidades de información.
- Modelos:
 - *clasificación de texto*
 - *extracción de texto.*



Nuestra motivación

¿Qué tanto se puede
analizar un chat?



Extracción de patrones

Estadísticas generales, del
comportamiento del grupo, de los
miembros.



Análisis de sentimientos

Emociones transmitidas, sentimientos
predominantes en los miembros o en
el chat.

Nuestro enfoque

Explorar las herramientas
disponibles en Python.



Objetivo # 1

Examinar las bibliotecas
para análisis de texto.



Objetivo # 2

Desarrollar un programa
para el análisis de chats.



Objetivo # 3

Analizar los resultados de
diferentes chats.

Camino a seguir



Exportar los chats.



Limpiar la data.



Extraer patrones.



Analizar sentimientos.

Librerías de Python

NLTK

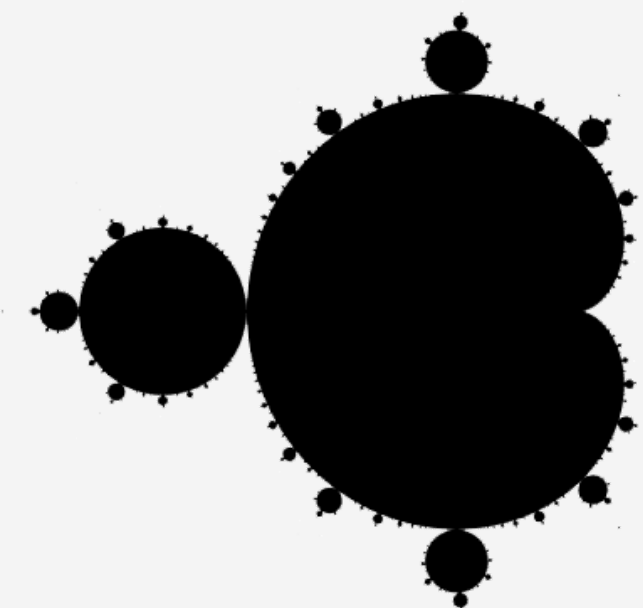
SentimentIntensityAnalyzer()
método: polarity_scores(texto)

TEXTBLOB

Textblob(texto)
método: sentiment()

DEEP_TRANSLATOR

GoogleTranslator...



TextBlob

¿Cómo se analizan los sentimientos?

Ciencia de Datos II

VADER

- Negativo (entre 0 y 1)
- Neutral (entre 0 y 1)
- Positivo (entre 0 y 1)
- Compuesto (entre -1 y 1)

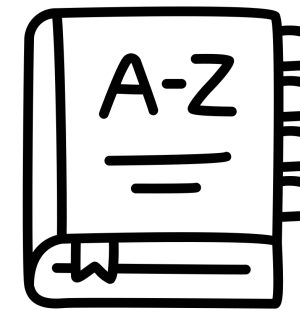
TEXTBLOB

- Polaridad (entre -1 y 1)
- Subjetividad (entre 0 y 1)
 - Adicionalmente, se calcula el complemento para obtener la objetividad.

Pasos internos:



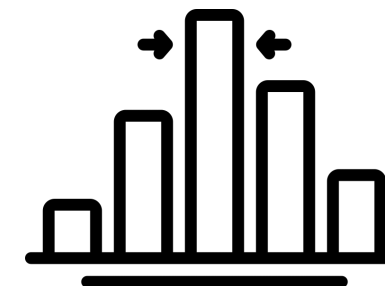
Separar el texto en palabras.



Consultar las palabras en sus diccionarios.



Sumar los pesos de cada palabra.



Normalizar la suma.

Planteamiento del programa

¿Qué lenguajes usar?



R (versión 4.4.0)

Para limpiar y transformar los datos "crudos".



Python (versión 3.11.7)

Análisis de texto y análisis de sentimientos.

Clase principal: ProcesadorTexto

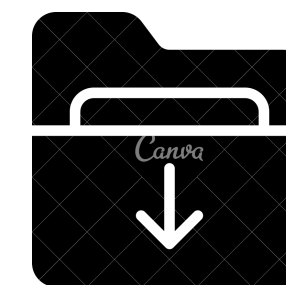
Métodos

Ciencia de Datos II



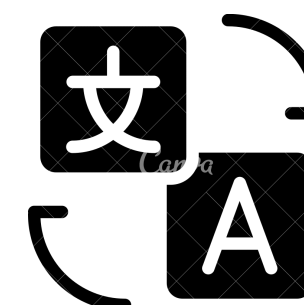
leer(nombre)

Lee un csv y los guarda en el DataFrame.



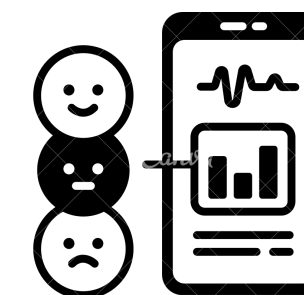
guardar(nombre)

Guarda el DataFrame en un csv.



traducir()

Traduce los mensajes de español a inglés.



analizar_sentimientos()

Realiza el análisis de sentimientos usando Vader y TextBlob.

Clase que hereda

AnalizadorTexto

Métodos

- ◆ contar_mensajes()
- ◆ contar_editados()
- ◆ encontrar_frase(frase)
- ◆ racha()
- ◆ dia_mas_concurrido()
- ◆ hora_promedio()

Clase que hereda

AnalizadorTexto

Métodos

- ◆ promedio_sentimientos()
- ◆ sentimiento_predominante(autor)
- ◆ autor_predominante(sentimiento)
- ◆ autores_destacados(medida)

Resultados: Gen

Ciencia de Datos II

TOP 10 MÁS ÁCTIVOS

Posición	Miembro	Cantidad	Posición	Miembro	Cantidad
1	Venegas	887	6	Dixon Montero	112
2	Javier Hernández	265	7	Juan Pablo M.	98
3	Anthony	144	8	Adrian	93
4	Corella	118	9	Luis Fer Amey	90
5	Ashley Padilla	114	10	Gustavo Amador	68

Fuente: elaboración propia.

Resultados: Gen

Ciencia de Datos II

TOP 5 MÁS EDITORES

Posición	Miembro	Cantidad
1	Javier Hernández	8
2	Venegas	5
3	Adrian	4
4	Ashley Padilla	3
5	Gustavo Amador	1

Fuente: elaboración propia.

Resultados: Gen

TOP 5 CON MÁS STICKERS ENVIADOS

Posición	Miembro	Cantidad
1	Venegas	88
2	Javier Hernández	32
3	Juan Pablo M.	25
4	Paula	25
5	Feliz	16

Fuente: elaboración propia.

Resultados: Gen

Ciencia de Datos II

TOP 5 CON MÁS IMÁGENES ENVIADAS

Posición	Miembro	Cantidad
1	Venegas	28
2	Javier Hernández	16
3	Sofía Bocker	12
4	Adrian	12
5	Dixon Montero	9

Fuente: elaboración propia.

Resultados: Gen

OTROS RESULTADOS

RACHA

- La mayor racha de días fue de 18, desde el 2023-11-26 al 2023-12-13

DÍA MÁS CONCURRIDO

- El día más concurrido es el 2024-05-22, con 308 mensajes

Resultados: Gen

Ciencia de Datos II

AUTOR PREDOMINANTE POR SENTIMIENTO

Sentimiento	Miembro	Valor
Positivo	Juan Pablo M.	0,2895
Negativo	Luis Fer Amey	0,1203
Neutral	Lau Villacís	1
Subjetivo	Adrian	0,2263
Objetivo	Juan Pablo M.	1

Fuente: elaboración propia.

Resultados: Herramientas II

Ciencia de Datos II

TOP 10 MÁS ÁCTIVOS

Posición	Miembro	Cantidad
1	 Potoy Juárez	141
2	Dixon	10
3	Gerard	8
4	Luis Fernando Amey	7
5	Marco Guardia	6

Posición	Miembro	Cantidad
6	Sofía Bocker	6
7	Corella	6
8	Paula	3
9	Montserrat Beirute A.	3
10	Santiago Actuariales	2

Fuente: elaboración propia.

Resultados: Herramientas II

Ciencia de Datos II

TOP 5 MÁS EDITORES

Posición	Miembro	Cantidad
1	Potoy Juárez	49
2	Sofía Bocker	3
3	Gerard	2
4	Luis Fernando Amey	1
5	Dixon	1

Fuente: elaboración propia.

Resultados: Herramientas II

Ciencia de Datos II

TOP CON MÁS STICKERS ENVIADOS

Posición	Miembro	Cantidad
1	Gerard	1
2	Paula	1

Fuente: elaboración propia.

Resultados: Herramientas II

Ciencia de Datos II

TOP 5 CON MÁS IMÁGENES ENVIADAS

Posición	Miembro	Cantidad
1	Potoy Juárez	13
2	Sofía Bocker	3
3	Montserrat Beirute A.	1
4	Luis Fernando Amey	1
5	Ashley Padilla	1

Fuente: elaboración propia.

Resultados: Herramientas II

OTROS RESULTADOS

RACHA

- La mayor racha de días fue de 6, desde el 2024-03-20 al 2024-03-25

DÍA MÁS CONCURRIDO

- El día más concurrido es el 2024-05-04, con 50 mensajes

Resultados: Herramientas II

Ciencia de Datos II

AUTOR PREDOMINANTE POR SENTIMIENTO

Sentimiento	Miembro	Valor
Positivo	dom	0,744
Negativo	Sofía Bocker	0,0672
Neutral	Adrian	1
Subjetivo	Javier Hernández	0.6
Objetivo	Adrian	1

Fuente: elaboración propia.

Resultados: Herramientas II

Ciencia de Datos II

AUTORES DESTACADOS (VARIABLE COMPUESTO)

Sentimiento	Miembro	Valor
Positivo	Javier Hernández	0,7777
Negativo	Adrian	0,0

Fuente: elaboación propia.

¿Qué sentimiento le predomina a Potoy?

Ciencia de Datos II

Sentimiento	Miembro	Valor
Positivo	Potoy Juárez	0,085

Fuente: elaboración propia.

Conclusiones y recomendaciones





CIENCIA DE DATOS II

Análisis de chats con Python

Anthony Mauricio Jiménez Navarro
Javier Antonio Hernández Navarro
Luis Fernando Amey Apuy