

Machine Learning

para

Análisis de Sentimiento

Una mirada en tiempo real al corazón de los consumidores con tecnologías de bajo coste al alcance de las pymes

En la era digital, la voz del consumidor resuena más fuerte que nunca. Para las pymes, comprender el sentimiento de sus clientes hacia su marca y productos ya no es un lujo, sino una necesidad competitiva que las nuevas tecnologías han puesto al alcance de sus limitados presupuestos. En este artículo, descubriremos - con un caso práctico en el que analizaremos los twitters sobre Iberia en la red social X (realizado íntegramente con Python), por qué, el análisis de sentimiento, una rama del procesamiento del lenguaje natural (NLP), se ha convertido en una herramienta indispensable para tomar el pulso del mercado en tiempo real y las muchas ventajas que puede aportar a tu negocio.

La revolución está aquí, y está al alcance de tu mano. ¿Estás listo para ser parte de ella?

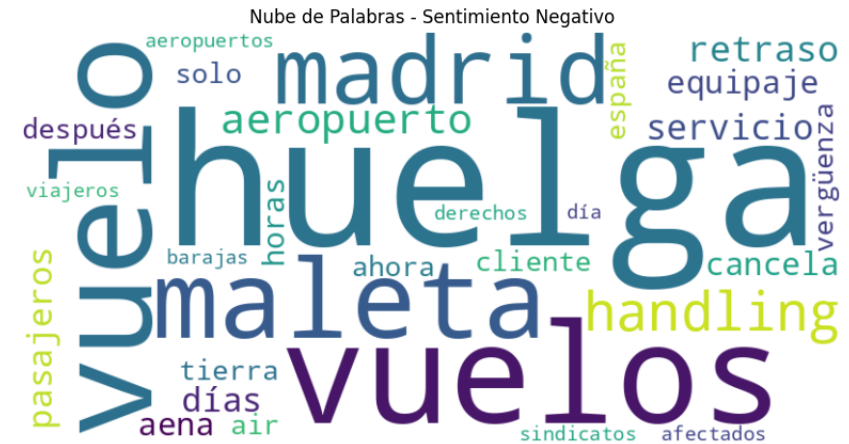
Puedes ver el código completo de Python en:

<https://github.com/EnriqueAranaz/AnalisisSentimiento>

Enrique Aranaz Tudela

Data Driven | Consultor BI | Data Scientist | Machine Learning Python | Microsoft Certified Power BI Data Analyst

Caso práctico



INDICE CONTENIDO

01 Introducción: El poder del análisis de sentimientos

02 Caso práctico: Iberia

2.1. Volando alto con datos: La importancia del análisis de sentimiento en X para Iberia

2.2. Implementación en Python y resultados obtenidos

03 Conclusiones



01 Introducción: El poder del análisis de sentimientos

En la era digital, la voz del consumidor resuena más fuerte que nunca. Para las pequeñas y medianas empresas comprender el sentimiento de sus clientes hacia su marca y productos ya no es un lujo, sino una necesidad competitiva. El análisis de sentimiento, una rama del procesamiento del lenguaje natural, se ha convertido en una herramienta indispensable para tomar el pulso del mercado en tiempo real.

¿Qué es el Análisis de Sentimiento?

El análisis de sentimiento es una técnica de procesamiento de lenguaje natural (NLP) que se utiliza para determinar la actitud o el sentimiento expresado en un fragmento de texto. Aquí tienes algunas aplicaciones comunes:

Monitoreo de Redes Sociales: Las empresas pueden analizar los comentarios en redes sociales para entender cómo se sienten los clientes sobre sus productos o servicios.

Feedback de Clientes: Analizar las opiniones y reseñas para identificar áreas de mejora y medir la satisfacción del cliente.

Análisis de Campañas: Evaluar la efectividad de las campañas publicitarias mediante el análisis de las reacciones del público.

Segmentación de Mercado: Identificar segmentos de clientes basados en sus sentimientos y opiniones.

Tendencias del Mercado: Detectar tendencias emergentes y cambios en las preferencias de los consumidores.

Competencia: Analizar los sentimientos hacia productos o servicios de la competencia.

Clima Laboral: Evaluar el sentimiento de los empleados a través de encuestas y comentarios para mejorar el ambiente de trabajo.

Selección de Personal: Analizar las respuestas de los candidatos en entrevistas para identificar actitudes y adecuación cultural.

Encuestas de Opinión: Analizar los resultados de encuestas para entender mejor las opiniones y sentimientos de la población.

Monitoreo de Crisis: Detectar y analizar rápidamente las reacciones públicas ante eventos críticos.

¿Por qué es crucial para las pymes?

1. Retroalimentación inmediata: Captura las opiniones de los clientes en el momento, permitiendo respuestas ágiles.

2. Mejora de productos: Identifica fortalezas y debilidades desde la perspectiva del consumidor.

3. Gestión de crisis: Detecta problemas potenciales antes de que se conviertan en crisis de reputación.

4. Ventaja competitiva: Proporciona insights valiosos para diferenciarse en el mercado.

5. Optimización de marketing: Ajusta estrategias basadas en la recepción real de campañas y mensajes.

La revolución Python: Democratizando el análisis de sentimiento

Afortunadamente, la barrera de entrada para implementar análisis de sentimiento se ha reducido significativamente gracias a Python y sus potentes librerías. Hoy, una pyme puede:

- Recopilar datos de redes sociales, reseñas y encuestas en tiempo real.
- Procesar grandes volúmenes de texto con eficiencia.
- Aplicar modelos de aprendizaje automático pre-entrenados para clasificar sentimientos.
- Visualizar resultados de forma clara y accionable.

Todo esto sin necesidad de un equipo de científicos de datos o una gran inversión en infraestructura. Python, con librerías como NLTK, TextBlob, y spaCy, ha democratizado estas capacidades, poniéndolas al alcance de empresas de todos los tamaños.

En este artículo, exploraremos - con un caso práctico analizando los twitters sobre Iberia en la red social X - cómo las pymes pueden aprovechar estas herramientas para sintonizar con la voz de sus clientes y transformar datos en decisiones estratégicas que impulsen su crecimiento.



02 Caso práctico: Iberia

Para ejemplificar la facilidad que supone el análisis de sentimientos en redes sociales he realizado un ejemplo práctico con Iberia en la red social X.

2.1. Volando alto con datos: La importancia del análisis de sentimiento en X para Iberia

En la era digital, las aerolíneas no solo compiten en los cielos, sino también en las redes sociales. Para Iberia, la principal aerolínea de España, el análisis de sentimiento de los comentarios en X (anteriormente Twitter) no es solo una herramienta de marketing, sino una brújula estratégica para navegar en un mercado altamente competitivo y sensible a la experiencia del cliente.

1. Pulso de la satisfacción del cliente en tiempo real

Proporciona una visión instantánea y continua de la satisfacción de sus clientes. Cada tuit es una pieza de retroalimentación directa, permitiendo a la aerolínea:

- Identificar problemas recurrentes en servicios específicos.
- Medir el impacto de nuevas iniciativas o cambios en las políticas.
- Detectar tendencias emergentes en las preferencias de los pasajeros

Esta información en tiempo real permite a Iberia ajustar rápidamente sus servicios, mejorando continuamente la experiencia del cliente.

2. Gestión de crisis y reputación

En la industria aérea, las crisis pueden desarrollarse rápidamente. El análisis de sentimiento en X ayuda a Iberia a:

- Detectar problemas potenciales antes de que se conviertan en crisis mayores.
- Medir el impacto de incidentes (como retrasos o cancelaciones) en la percepción de la marca.

- Adaptar su comunicación en tiempo real para abordar preocupaciones específicas de los clientes.

Esta capacidad de respuesta rápida es crucial para mantener la confianza de los clientes y proteger la reputación de la marca.

3. Inteligencia competitiva

Al analizar no solo los comentarios sobre Iberia, sino también sobre sus competidores, la aerolínea puede:

- Identificar áreas donde supera a la competencia y donde necesita mejorar.
- Descubrir oportunidades de mercado no atendidas.
- Anticipar movimientos de la competencia basados en el sentimiento de sus clientes.

Esta inteligencia competitiva es invaluable para la planificación estratégica a largo plazo.



4. Optimización de rutas y servicios

Iberia puede utilizar estos insights para optimizar su red de rutas, mejorar sus servicios a bordo y personalizar su oferta para diferentes tipos de viajeros.

5. Innovación impulsada por el cliente

El análisis de sentimiento puede ser una fuente rica de ideas para la innovación:

- Identificar nuevas necesidades o deseos expresados por los pasajeros.
- Evaluar la recepción de nuevas tecnologías o servicios implementados.
- Descubrir oportunidades para diferenciarse en el mercado.

Al escuchar atentamente a sus clientes en X, Iberia puede mantenerse a la vanguardia de la innovación en la industria aérea.

6. Mejora de la eficiencia operativa

Los comentarios en redes sociales pueden resaltar áreas de ineficiencia operativa:

- Problemas recurrentes en aeropuertos específicos.
- Cuellos de botella en procesos como el check-in o el embarque.
- Feedback sobre la tripulación y el personal de tierra

Esta información permite a Iberia abordar problemas operativos de manera proactiva, mejorando la eficiencia general de sus operaciones.

2.2. Implementación en Python y resultados obtenidos

1. Extracción de tweets de X

Aunque la forma más fácil es conectándose a la API de X, dado que para ello se necesita crear una cuenta de desarrollador y es de pago, he hecho uso de una de las muchas librerías gratuitas que posee Python que brindan acceso a infinidad de datos y tareas de extracción de información. En este caso, he utilizado Snsrape, herramienta que permite extraer datos de varias redes sociales sin necesidad de autenticación.

Extraemos aquellos tweets que contengan #iberia:

```
[ ] from ntscraper import Nitter

#Inicializar el scraper y obtener los tweets

# Inicializa el scraper
scraper = Nitter(0)

# Obtén tweets con límites de fechas
tweets = scraper.get_tweets("#iberia lang:es", mode='hashtag', number=100, until='2024-09-30')
```

Y los preprocesamos para eliminar la información que no nos es útil:

```
# Procesa los tweets
final_tweets = []
for x in tweets['tweets']:
    data = [x['link'], x['text'], x['date'], x['stats']['likes'], x['stats']['comments']]
    final_tweets.append(data)

# Crea el DataFrame
dat = pd.DataFrame(final_tweets, columns=['twitter_link', 'text', 'date', 'likes', 'comments'])

# Función para eliminar todo lo que venga después de y incluyendo '/status'
def remove_status(link):
    return link.split('/status')[0]

# Aplicar la función a la columna 'twitter_link'
dat['twitter_link'] = dat['twitter_link'].apply(remove_status)

# Guarda el DataFrame en un archivo Excel sin incluir el índice
dat.to_excel('tweets_iberia.xlsx', index=False)
```

Cómo resultado obtenemos el siguiente dataframe, con 2.259 tweets:

	twitter_link		text	date	likes	comments
0	https://twitter.com/novedadesaereas	⚠️⚠️	O'Leary menciona "ganas de que salga adelante la operación" de #Iberia y #AirEuropa y cree que un 30% de los slots de IB en #Madrid deberían c...	Apr 1, 2024 · 1:43 PM UTC	146	11
1	https://twitter.com/dinkyviajeros	🗋️	¿Tu vuelo ha sido cancelado y no sabes cómo reclamar el reembolso del billete? 😊 En este tutorial te explicamos cómo reclamar a la principales ...	Apr 1, 2024 · 2:30 PM UTC	0	1
2	https://twitter.com/romtolousse		Parece que ahora sí... al menos tenemos avion #coruña #iberia #vuelocancelado	Apr 1, 2024 · 4:06 PM UTC	3	0
3	https://twitter.com/AzucenaPelaez	@Iberia @consumidores @TurismoMadrid @minturgob @consumogob @AesaSpain	Tras varias conversaciones en privado con #IBERIA, siguen negando su respo...	Apr 1, 2024 · 6:18 AM UTC	0	2
4	https://twitter.com/conexiohispanoa	Iberia anuncia vuelos directos Madrid-Tokio desde octubre 2024, fortaleciendo conexiones entre España, Japón y América Latina. ✈️ ESJP #VuelosDirec...		Apr 1, 2024 · 7:31 PM UTC	0	0
...
2254	https://twitter.com/pdispain	El TSJM confirma el despido disciplinario de un trabajador de Iberia por sustraer una pantalla de ordenador https://postly.app/3cTR #empleo #iberi...		Sep 2, 2024 · 7:05 AM UTC	0	0
2255	https://twitter.com/raulamaros	Está situación de @iberia ya rebasa los límites de la paciencia, llevamos UNA semana SIN equipaje, son 4 maletas y me avisan que van a entregar UN...		Sep 2, 2024 · 5:38 AM UTC	0	1
2256	https://twitter.com/America_Vuela	Iberia lanza campaña de precios especiales para conectar México con Europa #AméricaVuela #AviaciónComercial #Aviación #Iberia #Campaña #PreciosEs...		Sep 1, 2024 · 10:00 PM UTC	1	0
2257	https://twitter.com/avclubcenter	#ActualidadAviacion Iberia aumentará su capacidad internacional desde el mes de septiembre Más información: https://aviationclubcenter.com/inde...		Sep 1, 2024 · 1:58 PM UTC	0	0
2258	https://twitter.com/MujerEjecutiva_	#Viajes ✈️ Iberia ha lanzado una campaña de descuentos para fortalecer la conexión entre México y Europa. Lee la nota completa aquí en nuestro...		Sep 1, 2024 · 12:01 AM UTC	0	0

2259 rows x 5 columns

Para finalizar con el preprocesamiento del dataframe, procederemos a eliminar los duplicados (51), y agrupar aquellos que siendo de distintas fechas, tienen el mismo texto (agrupamos en un solo tweet sumando los likes, comments, y quedandonos con la primera de las fechas).

```
# Chequear valores duplicados
print("---- Duplicados ----")
print("Número de duplicados: " + str(dat.duplicated().sum()))
print("")

# Eliminar duplicados
dat = dat.drop_duplicates()
```

---- Duplicados ----
Número de duplicados: 51

```
# Eliminar duplicados en texto y sumar 'likes' y 'comments'
dat_grouped = dat.groupby(['twitter_link', 'text']).agg({
    'date': 'first',
    'likes': 'sum',
    'comments': 'sum'
}).reset_index()
# dimension de los datos tras eliminar textos duplicados es
print("El tamaño del dataset tras eliminar textos duplicados es : ", dat_grouped.shape)
print("")
```

El tamaño del dataset tras eliminar textos duplicados es : (1942, 5)

Una vez tenemos los tweets convenientemente organizados llega el momento de asignarles una etiqueta en función del sentimiento, positivo, negativo o neutro que muestra cada uno de ellos. En este caso disponemos de dos metodologías posibles:

Método 1: Utilizando librerías disponibles

Las librerías de Python ofrecen una forma rápida y eficiente de implementar el análisis de sentimiento. Existen muchas librerías que nos van a permitir todo tipo de tareas relacionadas con el procesamiento de lenguaje natural (NLP). Entre las más destacadas:

- **TextBlob:** librería sencilla que proporciona una API para realizar varias tareas de procesamiento de lenguaje natural, incluyendo el análisis de sentimiento.
- **VADER (Valence Aware Dictionary and sEntiment Reasoner):** muy enfocada para sentimientos expresados en redes sociales.
- **BERT: Bidirectional Encoder Representations from Transformers:** desarrollado por Google, ha revolucionado el campo del procesamiento del lenguaje natural (NLP), incluyendo el análisis de sentimiento.

Método 2: Creando tu propio modelo

Crear tu propio modelo te da más control y puede ser más preciso para tu dominio específico, pero requiere de más trabajo, pues requieres de un conjunto de textos previamente etiquetados con los sentimientos que quieras evaluar (positivo, negativo, neutro, o cualquier variación) y entrenar el modelo usando algoritmos de aprendizaje automático como Naive Bayes, SVM, redes neuronales, etc.

Cómo ventaja añadir que crear tu propio modelo requiere más trabajo inicial, pero puede ofrecer mejores resultados a largo plazo, especialmente si tienes datos específicos de tu industria.



Para nuestro caso práctico me he inclinado por la primera opción, y utilizaré BETO, el modelo de BERT en español, entrenado utilizando un gran corpus de textos en español (libros, noticias, etc.).

BERT y BETO: Revolucionando el análisis de sentimiento

BERT: Bidirectional Encoder Representations from Transformers

BERT, desarrollado por Google, ha revolucionado el campo del procesamiento del lenguaje natural (NLP), incluyendo el análisis de sentimiento.

Características clave de BERT:

- 1. Bidireccionalidad:** BERT analiza el contexto de una palabra en ambas direcciones, lo que le permite entender mejor el significado en contexto.
- 2. Pre-entrenamiento:** Se entrena en un corpus masivo de texto no etiquetado, lo que le permite capturar patrones lingüísticos complejos.
- 3. Fine-tuning:** Puede ser afinado para tareas específicas, como el análisis de sentimiento, con relativamente pocos datos etiquetados.
- 4. Manejo de contexto:** Comprende matices y sutilezas del lenguaje mejor que modelos anteriores.

Características de BETO:

- 1. Especialización en español:** Entrenado en un gran corpus de texto en español, captura mejor las particularidades del idioma.
- 2. Variantes regionales:** Puede manejar diferentes variantes del español de distintas regiones.
- 3. Compatibilidad:** Utiliza la misma arquitectura que BERT, lo que permite aprovechar las herramientas y técnicas desarrolladas para BERT.

Aplicación del modelo Beto a nuestros datos

El modelo BETO es fácil de implementar, pues requiere tan sólo unas pocas líneas de código

```
from transformers import AutoTokenizer, AutoModelForSequenceClassification
import torch

#Cargar el modelo y el tokenizador de BETO
tokenizer = AutoTokenizer.from_pretrained("finiteautomata/beto-sentiment-analysis")
model = AutoModelForSequenceClassification.from_pretrained("finiteautomata/beto-sentiment-analysis")

#Definir la función de análisis de sentimiento
def analizar_sentimiento(texto):
    inputs = tokenizer(texto, return_tensors="pt")
    outputs = model(**inputs)
    logits = outputs.logits
    predicted_class_id = torch.argmax(logits).item()
    labels = ["NEG", "NEU", "POS"]
    return labels[predicted_class_id]

#Aplicar la función de análisis de sentimiento a cada tweet
dat_grouped['sentimiento'] = dat_grouped['text'].apply(analizar_sentimiento)
dat_grouped
```

Y ya tenemos nuestros datos con la característica “sentimiento” añadida a nuestro dataset:

	twitter_link	text	date	likes	comments	sentimiento
0	https://twitter.com/205winstonsmith	no ya compensar, pero por lo menos devolverme ...	Jan 21, 2024 · 5:21 PM UTC	0	1	NEG
1	https://twitter.com/21NoticiasCom	.@UGT_Galicia valora os avances na reunión con...	May 29, 2024 · 11:21 AM UTC	0	0	POS
2	https://twitter.com/7TeleValencia	✈️ HUELGA AÉREA: Iberia cancela 444 vuelos p...	Dec 30, 2023 · 9:00 AM UTC	0	0	NEG
3	https://twitter.com/AEMET_LaRioja	Iberia acusa a Aena de "irregularidades" en la...	Oct 23, 2023 · 9:25 AM UTC	0	0	NEG
4	https://twitter.com/AHeathrow	@Iberia llevo desde el jueves intentando compr...	Mar 11, 2024 · 11:27 AM UTC	0	7	NEG
...
1937	https://twitter.com/yenbaez	@Iberia @iberiaexpress Estimados desde el 26 d...	Dec 27, 2023 · 4:08 PM UTC	0	1	NEU
1938	https://twitter.com/yenbaez	El caso lleva desde el 26 de septiembre en trá...	Dec 29, 2023 · 7:31 PM UTC	0	2	NEG
1939	https://twitter.com/yolanda_sanz	#Iberia pierde tus maletas #iberiaplato, dos...	Jul 21, 2024 · 2:05 PM UTC	0	0	NEG
1940	https://twitter.com/yomismovaya	Putá mierda de iberia #iberia @Iberia_en	Jul 17, 2024 · 9:32 AM UTC	0	0	NEG
1941	https://twitter.com/zafirotoursofic	Hemos recogido en el stand de Amadeus un recon...	Jan 25, 2024 · 8:45 AM UTC	0	0	POS

Y este es el resultado sobre nuestros 2042 Tweets:

count	
sentimiento	
NEU	860
NEG	657
POS	525

Palabras más usadas en tweets calificados cómo negativos:

frequency			
huelga	230	españa	33
vuelo	126	tierra	33
vuelos	123	ahora	32
maleta	115	cliente	31
madrid	89	barajas	30
handling	80	vergüenza	30
aeropuerto	76	sindicatos	29
días	62	aeropuertos	29
equipaje	54	afectados	29
retraso	54	viajeros	28
servicio	51	día	28
pasajeros	48	hacer	28
después	42	derechos	28
aena	42	aerolínea	28
cancela	41	aireuropa	27
horas	37	hace	27
solo	36	hoy	26

Si nos fijamos en los Tweets que han sido calificados cómo negativos, vemos que el modelo parece haber captado muy bien el sentimiento :

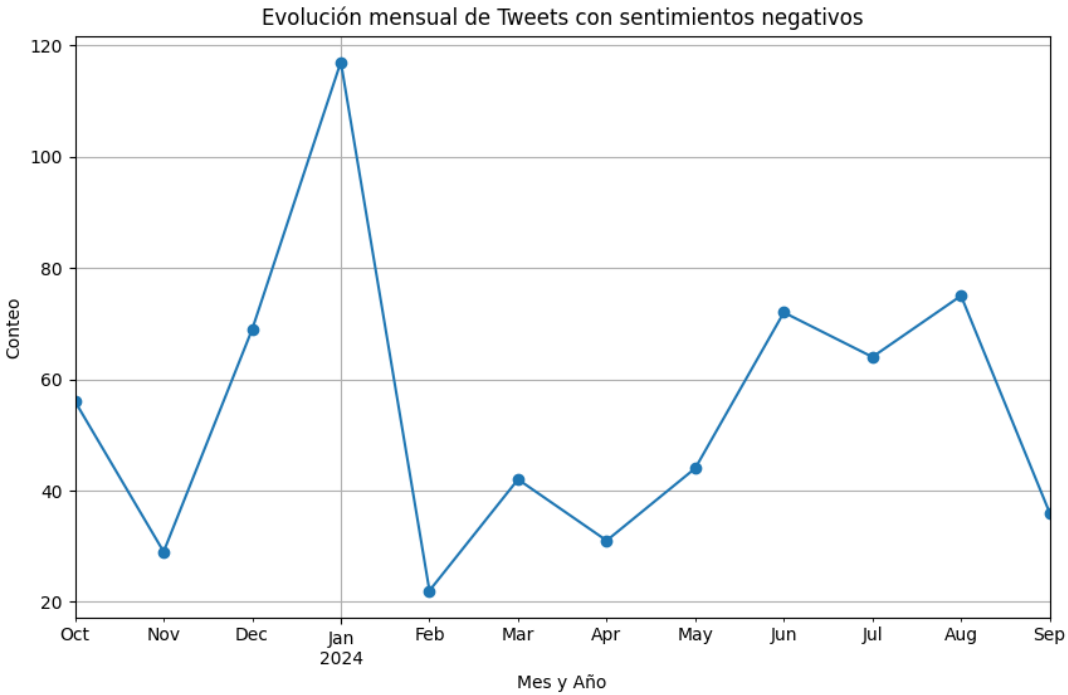
text	comments	likes	sentimiento
Otro día que pasa sin que @iberia me devuelva la plata de la reserva (ya emitida) que cancelaron. Seguro van a responder diciendo que me mandaron ...	9	0	NEG
Asi trata @Iberia a su gente. Después de cancelar el vuelo sin ninguna respuesta sus agentes, no esperan ni 5 minutos para cerrarte la conversaci...	7	0	NEG
@Iberia cuarta vez que nos perdéis una maleta. 20 horas de viaje y ahora llevamos 45 minutos haciendo cola para reclamar un equipaje que perdistei...	7	9	NEG
@Iberia Que vergüenza que traten el equipaje de mano así , la maleta llegó dañada , rota y nadie responde por los daños!! #iberia#vergüenza	7	1	NEG
Iberia #iberia ahora me. Atiende Evelyn Zambrano y me dice usted voló el 31 y ese vuelo ae canceló. Que burda respuesta teniendo yo toda la inform...	6	2	NEG
He presentado un reclamo en @AesaSpain ya que @Iberia no me compensa por un retraso de 9 hs en un vuelo de @FlyWithLEVEL . A la espera de cómo fun...	6	1	NEG
@Iberia que desastre retraso mal gestión del acceso a bordo. Todo los días mandado información pero del retraso no, hoy tb correo de la segunda co...	5	1	NEG
Increible @Iberia que te pagas un plus de asiento (con lo que ya cuestan de por sí los billetes) y al hacer el check-in te sientan donde les sale ...	5	0	NEG
Frustración no, indignación. los procedimientos ya se han seguido y habéis demostrado que os aprovecháis del débil y sois una empresa con muy poca...	5	3	NEG
@Iberia es normal que en una hora no hayan salido las maletas del vuelo de Málaga? Llevo más en barajas que en el aire! Aquí el ambiente se esta c...	5	1	NEG
#Iberia TIENE RESARCIRNOS POR EL RETRASO EL ALOJAMIENTO LA COMIDA Y LOS TICKTS DE TREN QUE PERDIMOS!!! Ya que todo sucedió dentro de la EU y rige ...	5	0	NEG
@Iberia @consumidores @consumomadrid @minturgob @consumogob @AesaSpain Historia de cómo #IBERIA arruina vacaciones y no se responsabiliza de NADA...	4	4	NEG
Otro retraso de @Iberia , otra vez que toca pagar cena aquí, que no tienes quien te busque y toca pagar taxi y otra vez que por supuesto no van a ...	4	2	NEG
#Cantabria avisa a #Iberia: la cancelación de #vuelos #Santander-#Madrid "perjudica" su imagen y las relaciones con Gobierno @cantabriaes @Iberia ...	3	2	NEG





Enlaces de Twitter con más tweets negativos

	Conteo	Porcentaje
twitter_link		
https://twitter.com/VincentOrtais	11	1.67
https://twitter.com/arturashe	11	1.67
https://twitter.com/sindicatostavla	11	1.67
https://twitter.com/RTVCes	9	1.37
https://twitter.com/JuanjoMorodo	8	1.22
https://twitter.com/JeannetteRen	7	1.07
https://twitter.com/Gacetinmadrid	6	0.91
https://twitter.com/Angelakutcher	6	0.91
https://twitter.com/RTVCCanarias	6	0.91
https://twitter.com/Antidiota1	5	0.76



03 Conclusiones

El procesamiento del lenguaje natural (NLP), se ha convertido en una herramienta indispensable para las pequeñas y medianas empresas en la era digital, permitiendo a las pymes comprender y responder eficazmente a las opiniones y necesidades de sus clientes, y ayudándoles a mejorar así su competitividad en el mercado.

Python, con sus numerosas librerías gratuitas y de código abierto, ha democratizado el acceso a estas técnicas avanzadas de análisis, permitiendo que incluso las empresas con recursos limitados puedan implementar soluciones sofisticadas de NLP.

La capacidad de realizar análisis de sentimiento de manera asequible y eficiente proporciona a las pymes insights valiosos para la toma de decisiones, la mejora de productos y servicios, y la gestión de la reputación de marca.

En un mundo donde la voz del cliente es cada vez más importante, el análisis de sentimiento se perfila como una ventaja competitiva crucial, permitiendo a las pymes no solo sobrevivir, sino prosperar en un entorno empresarial cada vez más data-driven y centrado en el cliente.

Puedes ver el código completo de Python en:
<https://github.com/EnriqueAranaz/AnalisisSentimiento>

Enrique Aranaz Tudela

*| Data Driven | Consultor BI | Microsoft Certified
Power BI Data Analyst | Data Scientist | Machine
Learning Python*



www.linkedin.com/in/enrique-aranaz-tudela



aranaz.enrique@gmail.com
aranaz.enrique@biparatupyme.com





Thank you