

Análisis de sentimientos de reseñas de sitios turísticos

MelA Reto 2023

Miguel Ángel Álvarez Carmona
Helena Gómez Adorno
Victor Giovanni Morales

CIMAT-MTY
IIMAS-UNAM
FCC-BUAP

miguel.alvarez@cimat.mx

26 de junio de 2023

Contenido

- 1 Introducción
- 2 Datos para la tarea de Análisis de sentimientos
- 3 ¿Cuál es la dinámica?
- 4 ¿Cómo mandar sus resultados?
- 5 ¿Cómo se van a evaluar los resultados?
- 6 Baseline

Contenido

- 1 Introducción
- 2 Datos para la tarea de Análisis de sentimientos
- 3 ¿Cuál es la dinámica?
- 4 ¿Cómo mandar sus resultados?
- 5 ¿Cómo se van a evaluar los resultados?
- 6 Baseline

- El turismo es un fenómeno social, cultural y económico relacionado con el desplazamiento de las personas a lugares fuera de su lugar habitual de residencia por razones personales o de negocios/profesionales.
- Esta actividad es vital en varios países, incluyendo México, donde el turismo representa el 8.7 % del PIB nacional, generando alrededor de 4.5 millones de empleos directos.
- Algunos países de LATAM se encuentra entre los diez primeros países a nivel mundial y de IberoAmérica en cuanto a la llegada de turistas internacionales.

- El Rest-Mex 2023 es un foro de evaluación dedicado al análisis de sentimientos en textos turísticos mexicanos.
- La tarea de análisis de sentimientos en textos turísticos ha ganado relevancia en los últimos años.
- Sin embargo, la mayoría de los esfuerzos científicos se han centrado en el idioma inglés, mientras que los textos en español, especialmente aquellos que no son de España, han recibido menos atención.
- Rest-Mex busca abordar esta brecha y enfocarse en el análisis de sentimientos en textos turísticos mexicanos.

Contenido

- 1 Introducción
- 2 Datos para la tarea de Análisis de sentimientos
- 3 ¿Cuál es la dinámica?
- 4 ¿Cómo mandar sus resultados?
- 5 ¿Cómo se van a evaluar los resultados?
- 6 Baseline

- El corpus del Rest-Mex 2023 es una recopilación de opiniones de turistas sobre lugares, restaurantes y hoteles en México, Cuba y Colombia.
- Este corpus se ha obtenido a partir de los comentarios compartidos en TripAdvisor durante un período de 20 años, desde 2002 hasta 2022.
- El corpus contiene más de 300,000 opiniones, lo que lo convierte en una fuente de datos rica y diversa para el análisis de sentimientos en el ámbito turístico.

Características del Corpus

- El corpus del Rest-Mex 2023 consta de opiniones etiquetadas con clases de polaridad que van del 1 al 5.
- Cada opinión representa el sentimiento expresado por un turista sobre su experiencia en un destino turístico específico.
- Las opiniones abarcan una amplia variedad de lugares, restaurantes y hoteles en México, Cuba y Colombia, lo que proporciona una visión completa de la satisfacción y las experiencias de los turistas en estos países.
- La difuclutad de este corpus recae en su desbalance de clases.

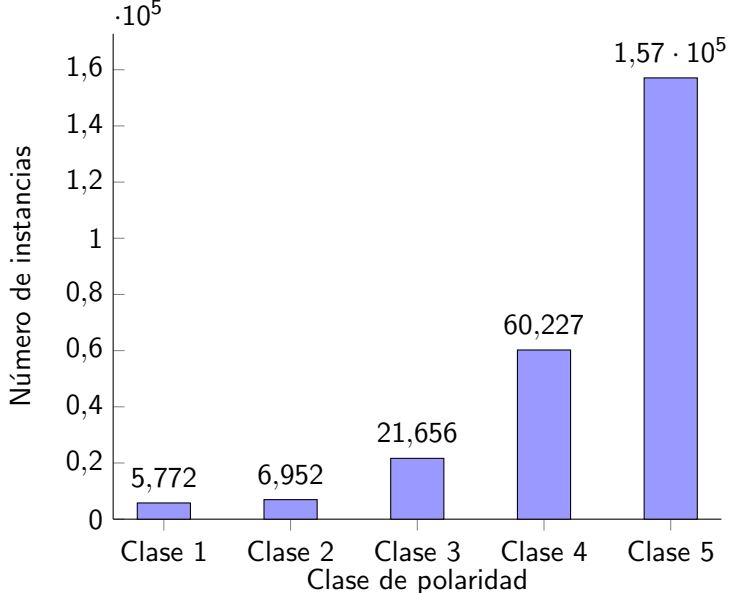


Figura: Comparación de las clases de polaridad y el número de instancias

Sub corpus para nuestro reto

- Para el corpus del Reto del MeIA se tomaron aleatoriamente 25 mil instancias
- Se trató de construir una distribución que mantenga el desbalance.
- Se tomará el 70 % de los datos para entrenamiento y el 30 % para la evaluación.

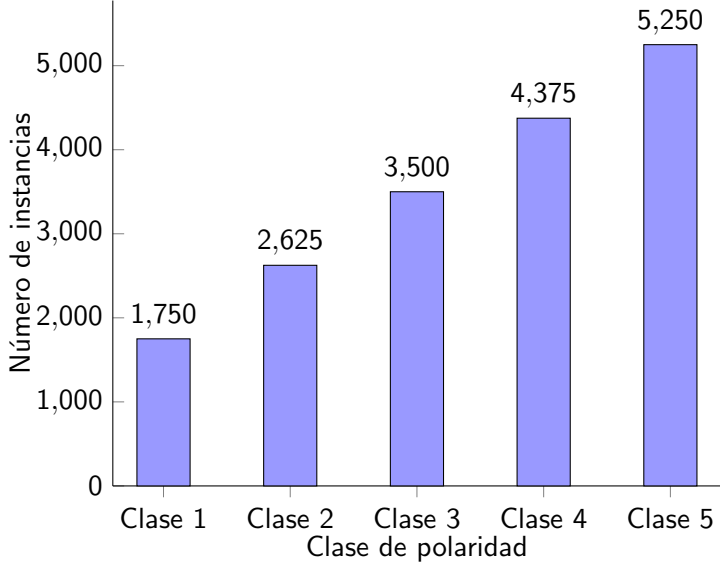


Figura: Comparación de las clases de polaridad y el número de instancias

Contenido

- 1 Introducción
- 2 Datos para la tarea de Análisis de sentimientos
- 3 ¿Cuál es la dinámica?
- 4 ¿Cómo mandar sus resultados?
- 5 ¿Cómo se van a evaluar los resultados?
- 6 Baseline

¿Cuál es la dinámica?

- Para la participación de este reto se seguirá el siguiente camino:

¿Cuál es la dinámica?

- Para la participación de este reto se seguirá el siguiente camino:
- A cada equipo se le compartirá el conjunto de entrenamiento etiquetado.

¿Cuál es la dinámica?

- Para la participación de este reto se seguirá el siguiente camino:
- A cada equipo se le compartirá el conjunto de entrenamiento etiquetado.
- Tambipen se les compartirá el conjunto de prueba sin etiquetas.

¿Cuál es la dinámica?

- Para la participación de este reto se seguirá el siguiente camino:
- A cada equipo se le compartirá el conjunto de entrenamiento etiquetado.
- Tambipen se les compartirá el conjunto de prueba sin etiquetas.
- Los participantes deben construir un modelo de clasificación para predecir las etiquetas del conjunto de prueba.
- Los participantes deben mandar su solución a los organizadores.
- Se hará la evaluación de sus resultados.

Contenido

- 1 Introducción
- 2 Datos para la tarea de Análisis de sentimientos
- 3 ¿Cuál es la dinámica?
- 4 ¿Cómo mandar sus resultados?
- 5 ¿Cómo se van a evaluar los resultados?
- 6 Baseline

Vamos a adivinar este ejemplo

Ejemplo

"Definitivamente un lugar para NO volver lastimosamente me lleve una muy mala imagen del lugar por infinidad de cosas. Los meseros que atendieron mi mesa no tenían idea de los platos que ofrecían en la carta y ni idea de atención al cliente..."

- ¿Qué etiqueta sería?

Vamos a adivinar este ejemplo

Ejemplo

"Definitivamente un lugar para NO volver lastimosamente me lleve una muy mala imagen del lugar por infinidad de cosas. Los meseros que atendieron mi mesa no tenían idea de los platos que ofrecían en la carta y ni idea de atención al cliente..."

- ¿Qué etiqueta sería?
- 1 (La más negativa)

¿Cómo mandar los resultados?

ID	Review
	Definitivamente un lugar para NO volver lastimosamente me lleve una muy mala imagen del lugar por infinidad de co
0	1. Los meseros que atendieron mi mesa no tenían idea de los platos que ofrecían en la carta y ni idea de atención al clien
1	Bueno es un. Hotel que ya es necesario un mantenimiento y una modernización urgente porque con el nivel que esta pri
2	Hoy fui a comer con mi novio a este lugar, pedimos enchiladas, pero tres enchiladas estaban crudas, le dije a la mesera,
	Estuve en la noche del 22 de sepbre.
3	Pésima atención
4	Pues para empezar te mandan lo que ellos quieren, especifique cómo querían mi servicio y me mandaron todo lo que le
5	Hola, el 29 de julio me hospedé en el Safi Valle en los cuartos 1215 y 1216. La mucama extrajo una bolsa de compras sir
6	Hola, les cuento q he hemos estado hospedados por tercera vez en este hotel y desgraciadamente como todas las cosas
7	Cuidado!!!! Hacen un cambio de moneda abusivo 16 90 pesos en la calle 120. la recomendación fue al plato mas caro. mi



My-Team: Bloc de notas				
Archivo	Edición	Formato	Ver	Ayuda
0	1			
1	5			
2	3			
3	3			
4	1			
5	4			
6	5			
7	2			
.	.			
.	.			
.	..			

Contenido

- 1 Introducción
- 2 Datos para la tarea de Análisis de sentimientos
- 3 ¿Cuál es la dinámica?
- 4 ¿Cómo mandar sus resultados?
- 5 ¿Cómo se van a evaluar los resultados?
- 6 Baseline

¿Cómo se van a evaluar los resultados?

$$Eval_P(k) = \frac{\sum_{i=1}^{|C|} ((1 - \frac{T_{C_i}}{T_C}) * F_i(k))}{\sum_{i=1}^{|C|} 1 - \frac{T_{C_i}}{T_C}}$$

- Donde k es un sistema participante del reto
- $C = \{1, 2, 3, 4, 5\}$
- T_C es el total de instancias en la colección
- T_{C_i} es el total de instancias en la clase i .
- Finalmente, $F_i(k)$ es el valor de la medida F para la clase i obtenido por el sistema k .

Contenido

- 1 Introducción
- 2 Datos para la tarea de Análisis de sentimientos
- 3 ¿Cuál es la dinámica?
- 4 ¿Cómo mandar sus resultados?
- 5 ¿Cómo se van a evaluar los resultados?
- 6 Baseline**

distilbert-base-multilingual-cased

Se aplicó un ajuste fino para la clasificación de texto sobre un modelo enmascarado llamado distilbert^a que es una versión destilada de bert-base-multilingual-cased. Los resultados obtenidos son en la primera época un F1 igual a **0.5121** y en la época dos un F1 de **0.5444**.

^a<https://huggingface.co/distilbert-base-multilingual-cased>

¡Gracias!

Análisis de sentimientos de reseñas de sitios turísticos

MeIA Reto 2023

Miguel Ángel Álvarez Carmona
Helena Gómez Adorno
Victor Giovanni Morales

CIMAT-MTY
IIMAS-UNAM
FCC-BUAP

miguel.alvarez@cimat.mx

26 de junio de 2023