



POLITÉCNICA

"Ingeniamos el futuro"

CAMPUS
DE EXCELENCIA
INTERNACIONAL



Graduado en Ingeniería Informática
Universidad Politécnica de Madrid
Escuela Técnica Superior de
Ingenieros Informáticos

Práctica NLP

Sistemas Inteligentes

Autor/a: M^a Yanira Calero Pereira

MADRID, Enero 2022

Índice

Introducción y objetivo del conjunto de datos elegido	2
Desarrollo del análisis	2
Referencias	4

Introducción y objetivo del conjunto de datos elegido

El principal objetivo de este trabajo es poder realizar un análisis de sentimientos según la clasificación que se haya hecho de los comentarios/valoraciones contenidas en el conjunto de datos seleccionado. En este conjunto de datos se recogen las distintas valoraciones que han realizado determinadas personas a diferentes restaurantes de Europa a través de la famosa aplicación de TripAdvisor, y se compone de 125527 observaciones y 11 variables. Entre los datos que se pueden visualizar estarían el nombre del país donde se encuentra el restaurante, su rango de precios, la media de puntuación (entre 1 y 5 estrellas), el número de reseñas que ha recibido, etc.

Desarrollo del análisis

En primer lugar, hemos obtenido los datos de los campos más relevantes del conjunto de datos elegido: *Name*, *Reviews*, *Ranking* y *Rating*. Una vez hecho esto, se ha realizado un análisis de los sentimientos agrupando las palabras para poder obtener cuáles han sido las que mayor aparición (repetición) tienen en las reviews (valoraciones) de los restaurantes del conjunto de datos. En este caso, se muestran tanto las palabras positivas como las negativas más empleadas:

	word	mean_rating	score	count_word
1	nice		3	13211
2	perfect	4.22993981945837	3	5982
3	love	4.21226927252986	3	4605
4	stop	3.9338146811071	-1	3324
5	bad	3.40299401197605	-3	1670
6	avoid	2.97113594040968	-1	1611

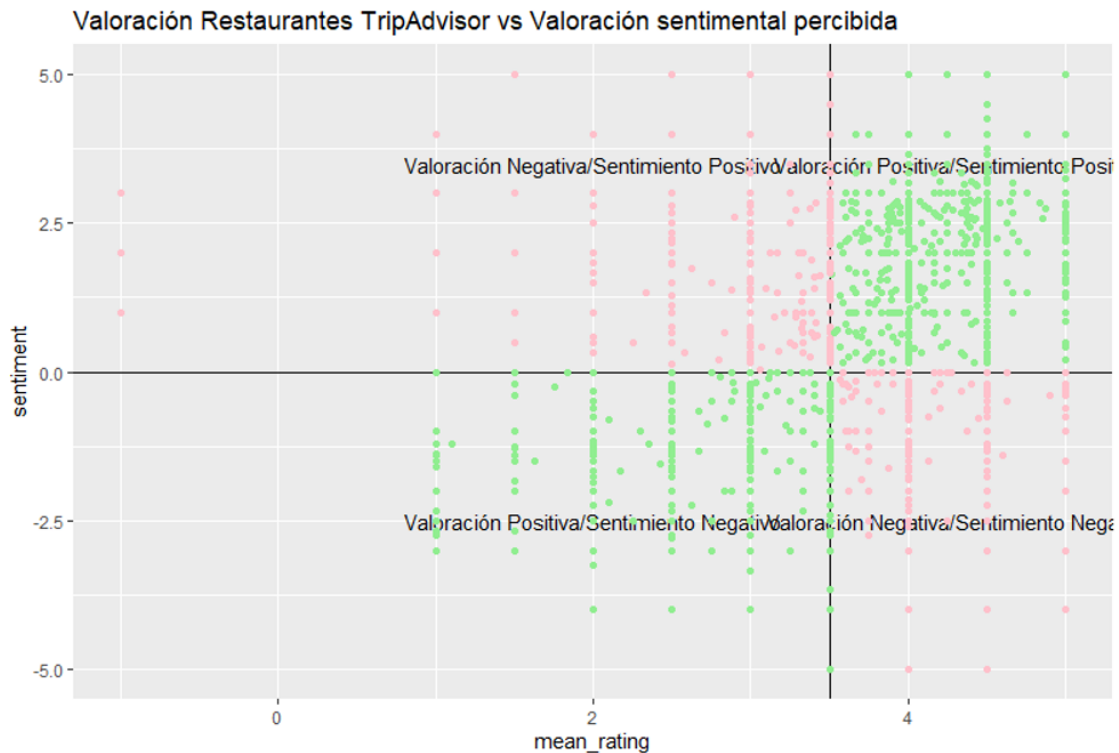
En la nube de palabras obtenida con la función *wordcloud* se pueden apreciar mejor gráficamente, siendo las palabras con destacadas con mayor tamaño las más usadas en las valoraciones de los clientes a los restaurantes:



De esta manera, en segundo lugar, se ha analizado qué palabras positivas han sido las más habituales en las reviews, siendo las más visibles las de mayor nivel de aparición en las valoraciones. En este caso, las palabras con connotaciones positivas más frecuentes en los comentarios han sido: *perfect*, *love*, *stop* y *fresh*:

[illegible]

Por último, en la siguiente gráfica se ha querido plasmar la relación entre los sentimientos positivos/negativos que se transmiten en función de las valoraciones positivas/negativas registradas en TripAdvisor. Por lo que se puede interpretar en la gráfica, vemos que algunas de las palabras negativas empleadas en las valoraciones pueden transmitir un sentimiento positivo, y, por el contrario, palabras positivas empleadas en las valoraciones pueden transmitir un sentimiento negativo. Con esto, podríamos decir que la relación entre las valoraciones de los clientes y los sentimientos que transmiten las palabras incluidas en ellos, en ocasiones, no llega a ser totalmente fiable.



Referencias

Dataset obtenido de Kagle	https://www.kaggle.com/damienbeneschi/krakow-ta-restaurans-data-raw
Moodle de la asignatura	https://moodle.upm.es/titulaciones/oficiales/course/view.php?id=3365