
TEXT MINING & IMAGE RECOGNITION
HOJA DE TRABAJO # 4

Instrucciones: A continuación verá una lista de ejercicios que debe completar para poder entregar el la hoja de trabajo #4. Al finalizar, todos sus archivos deben estar contenidos en un archivo ht4-sucarnet.zip. Este archivo lo debe entregar en el link del GES. Por favor cree una carpeta para cada ejercicio que usted realice.

Problema 1:

Para este problema deberá hacer análisis de sentimientos para un conjunto de reviews de Yelp, deberá mostrar las siguientes métricas sobre el conjunto de reviews:

1. Mostrar el porcentaje de comentarios positivos y comentarios negativos en una misma gráfica
2. Mostrar una gráfica tipo boxplot para la distribución de los resultados de la polaridad y la subjetividad
3. Deberá realizar un análisis estadístico con comentarios sobre los resultados de sus hallazgos a partir de las gráficas anteriores.

Problema 2:

Cree un web scraper para los reviews del sitio Trip Advisor, similar al demo trabajado durante la clase #8. Desarrolle un programa que le permita aplicar el análisis de sentimientos desarrollado por su persona en el ejercicio anterior, deberá analizar el código HTML para poder realizar exitosamente el web scraper.

Para el caso de Trip Advisor, podrá notar que la URL se modifica por medio de un parámetro llamado *or-X* donde X es el índice inicial de cada review. En Trip Advisor, los reviews están organizados en múltiplos de 5. A continuación se le muestra una URL y el parámetro de configuración mencionado:

`https://www.tripadvisor.com.mx/Hotel_Review-g662606-d283869-Reviews-or5-Jardines_de_Nivaria_Adrian_Hoteles-Costa_Adeje_Adeje_Tenerife_Canary_Islands.html#REVIEWS`