

# **PRÁCTICA**

# **Análisis de Sentimientos**

**Sistemas de Gestión Empresarial**

**Curso 2023/2024**

Profesor : Álvaro Juan Ciriaco

Fecha

## **INTRODUCCIÓN**

El análisis de sentimientos implica el **proceso de discernir el matiz emocional presente en una serie de expresiones escritas o habladas**. La herramienta de análisis de sentimiento constituye una técnica automatizada que busca extraer información valiosa de los clientes en relación a sus actitudes, emociones y opiniones.

Esta evaluación de sentimientos implica analizar las emociones, actitudes y opiniones expresadas en el contenido. Las empresas emplean este enfoque para adquirir perspicacia sobre cómo los clientes reaccionan a un producto o servicio específico.

Para llevar a cabo este proceso, la herramienta de análisis de sentimiento hace uso de tecnologías de inteligencia artificial, tales como el procesamiento del lenguaje natural, análisis de texto y ciencia de datos, con el propósito de identificar, extraer y examinar la información de naturaleza subjetiva. En términos más sencillos, esta herramienta clasifica el texto en categorías de positivo, negativo o neutral. Mientras que las métricas convencionales, como la cantidad de vistas, clics, me gusta, compartidos, comentarios, entre otros, se centran en aspectos cuantitativos, el análisis de sentimiento trasciende los números y pone su enfoque en la calidad de las interacciones entre el público y la organización.

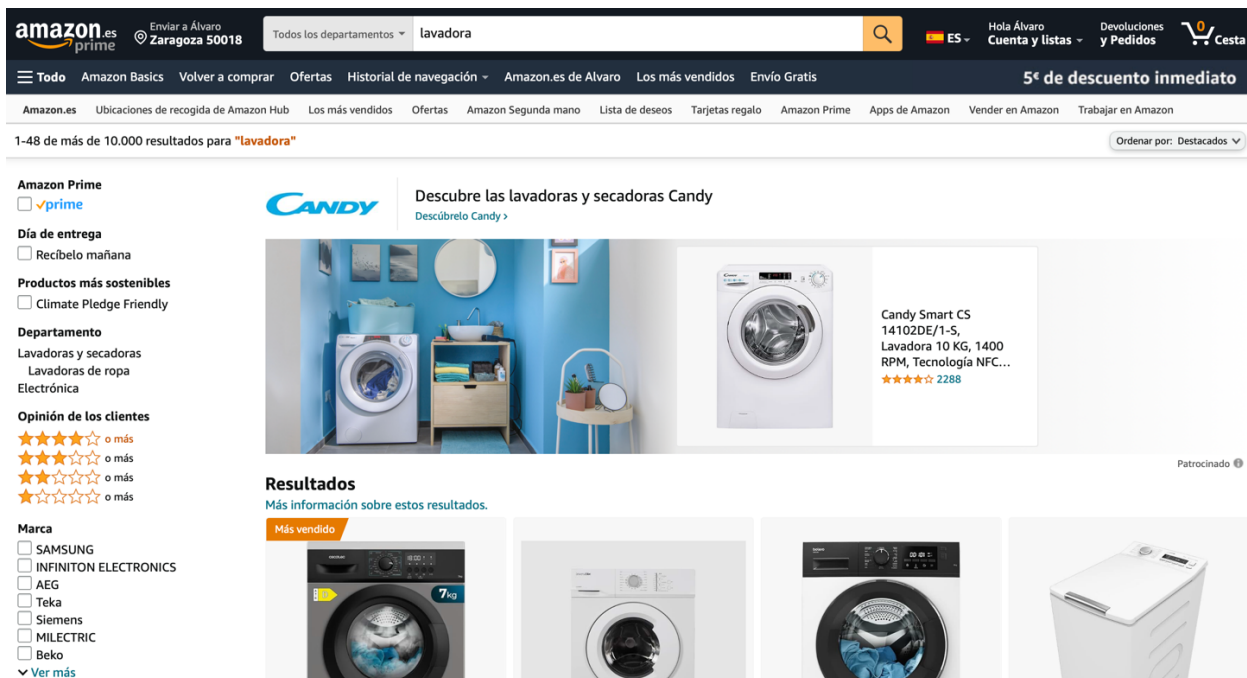
## **PARTE 1 : EXTRACCIÓN Y LIMPIEZA DE LA INFORMACIÓN NECESARIA**

La extracción de información se enmarca en **la recuperación de datos y tiene como objetivo principal la obtención automatizada de datos estructurados o parcialmente estructurados a partir de documentos legibles por los ordenadores**.

Un uso común de la extracción de información es la digitalización de una serie de documentos escritos en lenguaje natural, para luego almacenar la información extraída en una base de datos. Estos documentos pueden variar ampliamente en su estructura, desde textos parcialmente estructurados hasta completamente desestructurados. Pueden abarcar desde artículos periodísticos hasta informes científicos, generalmente escritos en lenguaje humano. Las tendencias actuales en extracción de información hacen uso de técnicas avanzadas de procesamiento del lenguaje natural, con un enfoque en dominios de aplicación específicos.

El proceso implica el uso de software de procesamiento de lenguaje natural (NLP) para analizar estos documentos y extraer información relevante. Sin embargo, esta tarea es compleja debido a que dichos programas tienden a funcionar en dominios muy limitados, lo que dificulta la extracción de información a partir de textos poco formales o incluso imágenes.

En el contexto de esta práctica, el objetivo del estudiante consiste en extraer contenido y valoraciones de al menos 10,000 comentarios de productos en Amazon, preferiblemente de productos similares. A continuación, se muestra un ejemplo de búsqueda centrada en lavadoras:



amazon.es prime Enviar a Álvaro Zaragoza 50018 Todos los departamentos lavadora

5€ de descuento inmediato

Amazon.es Ubicaciones de recogida de Amazon Hub Los más vendidos Ofertas Amazon Segunda mano Lista de deseos Tarjetas regalo Amazon Prime Apps de Amazon Vender en Amazon Trabajar en Amazon

1-48 de más de 10.000 resultados para "lavadora"

Amazon Prime

✓prime

Día de entrega

☐ Recíbelo mañana

Productos más sostenibles

☐ Climate Pledge Friendly

Departamento

Lavadoras y secadoras

Lavadoras de ropa

Electrónica

Opinión de los clientes

★★★★☆ o más

★★★★☆ o más

★★★★☆ o más

★★★★☆ o más

Marca

☐ SAMSUNG

☐ INFINITON ELECTRONICS

☐ AEG

☐ Teka

☐ Siemens

☐ MILETRIC

☐ Beko

Ver más

CANDY Descubre las lavadoras y secadoras Candy Descúbrelas Candy >

Candy Smart CS 14102DE/1-S, Lavadora 10 KG, 1400 RPM, Tecnología NFC... ★★★★★ 2288

Resultados

Más información sobre estos resultados.

Más vendido

Después de realizar una búsqueda en Amazon, el sistema muestra todos los productos que coinciden con las palabras clave seleccionadas por el usuario. Si se selecciona uno de estos productos, Amazon carga de manera completa toda la información relacionada con el mismo, incluyendo imágenes, precio y otros detalles relevantes. Además, al final de la página del producto, se presentan los comentarios de los usuarios que han comprado y utilizado el producto.

De acuerdo a los requisitos especificados, el estudiante debe crear un script en Python que extraiga información de 10,000 comentarios de productos en Amazon. A continuación, se proporciona un resumen del procedimiento:

1. **Entrada de Palabra Clave:** El script debe solicitar al usuario una palabra clave de búsqueda, por ejemplo, "lavadora".
2. **Búsqueda y Extracción de Datos:** El programa debe automatizar la búsqueda en Amazon y recopilar la información de productos necesaria para alcanzar al menos 10,000 comentarios. Esto incluirá la valoración (1 a 5 estrellas) y el texto asociado a cada comentario.
3. **Procesamiento de Datos:** El script debe limpiar los comentarios, eliminando caracteres irrelevantes que puedan interferir con el entrenamiento. Además, debe reducir las valoraciones de estrellas a dos opciones (0 para comentarios negativos y 1 para comentarios positivos).
4. **Almacenamiento en un archivo CSV:** Una vez que se han procesado los comentarios, el script debe guardar la información en un archivo CSV con dos columnas. La primera columna contendrá las valoraciones (0 o 1) y la segunda columna contendrá el texto procesado de la reseña.

Para llevar a cabo este proyecto, se pueden utilizar bibliotecas como **BeautifulSoup** para el web scraping de Amazon, y herramientas de procesamiento de texto en Python, como **nlTK** o **re**, para la limpieza de texto. Para la reducción de valoraciones de estrellas, puedes utilizar una función que asigne un valor de 0 o 1 según un umbral predefinido.

## **ENTREGA**

El trabajo debe ser realizado de forma individual y entregado dentro de los plazos establecidos en la tarea de Classroom. La entrega debe incluir el script en Python que cumple con las funciones requeridas en la tarea, junto con un documento de texto explicativo que consta de 3 páginas (que pueden contener imágenes). En este documento, el estudiante debe proporcionar instrucciones detalladas sobre el enfoque que ha seguido para completar la tarea.

Es importante destacar que, de acuerdo con el FIM, cualquier forma de copia o plagio en el trabajo resultará en la calificación automática de "suspense" para el trimestre correspondiente.