

## Taller 1: Extracción de Datos y Análisis de Sentimiento en E-commerce

### Integrantes del grupo

- OSCAR EDUARDO AVILA BERNAL
- JOSE DARIO NAVAS TORRES
- JUAN DAVID ORTEGA MURCIA
- NICOLAS ESTEBAN ROMERO VARGAS

### Introducción:

En este proyecto, se desarrolló una solución automatizada de extracción de datos y análisis de sentimiento aplicada al contexto de e-commerce. Utilizando técnicas de web scraping, se recolectaron datos clave de productos, como nombres, precios, calificaciones y comentarios de usuarios, con el fin de generar insights útiles para los consumidores y las empresas.

Además, se implementó un análisis de sentimiento para interpretar la percepción de los usuarios sobre los productos y una fórmula ponderada que calcula un puntaje de reputación integral. Este puntaje combina criterios como popularidad, calidad percibida y fiabilidad del vendedor, permitiendo así a los usuarios tomar decisiones más informadas.

## 1. Proceso de Extracción y Herramientas Utilizadas

### Herramientas

- **Requests:** Utilizada para realizar solicitudes HTTP y obtener el contenido de la página web.
- **BeautifulSoup:** Ayudó a analizar y extraer datos de las estructuras HTML de la página.
- **Pandas:** Para almacenar y manipular los datos extraídos en un formato tabular.
- **TextBlob:** Para realizar el análisis de sentimiento de los comentarios obtenidos.
- **GoogleTrans:** Traductor para convertir comentarios en otros idiomas al español, mejorando la consistencia del análisis.
- **NLTK:** Para tokenización, lematización y preparación del texto.

### Proceso

- I. **Identificación de la Página:** Seleccionamos Mercado Libre como fuente de datos debido a su amplio catálogo de productos y su relevancia en e-commerce.
- II. **Web Scraping Ético:** El script implementado respeta los términos de uso de Mercado Libre, evitando sobrecargar los servidores con solicitudes frecuentes.
- III. **Extracción de Datos:** Se extrajeron:
  - Nombre del producto
  - Precio
  - Características
  - Vendedor y su calificación
  - Comentarios de usuarios
- IV. **Transformación:** Los datos se almacenaron en un archivo CSV para facilitar su análisis posterior.

## 2. Resultados del Análisis de Sentimiento

## Taller 1: Extracción de Datos y Análisis de Sentimiento en E-commerce

### Metodología:

Se utilizó la biblioteca **TextBlob** para evaluar el sentimiento de los comentarios en términos de polaridad y subjetividad.

- La polaridad varía entre -1 (negativo) y 1 (positivo).
- La subjetividad refleja si un comentario es más objetivo (cercano a 0) o subjetivo (cercano a 1).
- Resumen de Sentimientos:
  - Comentarios positivos: 70%
  - Comentarios negativos: 20%
  - Comentarios neutrales: 10%
  - Promedio de subjetividad: 0.65

### 3. Observaciones Relevantes

#### Calidad de los Datos

- Los datos extraídos fueron en su mayoría consistentes, aunque algunos comentarios eran cortos y carecían de contexto.
- La calificación de los vendedores es un indicador clave para determinar la fiabilidad del producto.

#### Sesgos Potenciales

- Los comentarios traducidos pueden introducir imprecisiones en el análisis de sentimientos debido a limitaciones en las traducciones automáticas.
- La muestra de comentarios puede no representar el sentimiento general si solo se incluyen los más recientes.
- La página tiene restricciones en la cantidad de comentarios a los que se pueden acceder desde la librería **BeautifulSoup**
- El resultado de análisis de sentimientos y el puntaje final puede verse afectado por el funcionamiento interno de la compañía

### 4. Cálculo del Puntaje de Reputación

Se implementó una fórmula ponderada para calcular un puntaje de reputación para los productos. Este puntaje integra múltiples factores clave que determinan la calidad y confiabilidad del producto:

**Propósito:** Este puntaje ayuda a los usuarios a tomar decisiones informadas al comparar productos. Combina criterios como popularidad, calidad percibida, fiabilidad del vendedor, y análisis de sentimientos de los comentarios, otorgando mayor peso a los aspectos que el usuario considere más importantes mediante los parámetros de ponderación

$$Puntaje = 10 \times \left( w_1 \times \left( \frac{Cantidad\ Ventas}{max\_ventas} \right) + w_2 \times \left( \frac{Calificación\ Producto}{5} \right) + w_3 \times \left( \frac{Cantidad\ Opiniones}{max\_opiniones} \right) + w_4 \times \left( \frac{Rating\ Vendedor}{5} \right) + w_5 \times \left( \frac{Polaridad\ Final + 1}{2} \right) - w_6 \times Subjetividad\ Final \right)$$

### Conclusión

El análisis proporcionó una visión general de los productos y su recepción en la plataforma. Los resultados pueden servir como base para estrategias de marketing o decisiones de inventario en una organización de comercio electrónico.