

MercadoLibre Scraper & Análisis de Sentimientos

Una herramienta completa para extraer productos de MercadoLibre Argentina, obtener comentarios de usuarios y realizar análisis de sentimientos automatizado usando Python y Streamlit.

Tabla de Contenidos

- [Características](#)
- [Instalación](#)
- [Uso](#)
- [Configuración](#)
- [Análisis de Sentimientos](#)
- [Estructura de Datos](#)
- [Ejemplos](#)
- [Consideraciones](#)
- [Contribución](#)
- [Licencia](#)

Características

Web Scraping Avanzado

- Extracción de productos de MercadoLibre Argentina
- Soporte para múltiples páginas (hasta 5 páginas)
- Extracción de hasta **250 productos** por sesión
- Obtención de comentarios y calificaciones de usuarios

Análisis de Sentimientos

- Análisis automático de comentarios en español
- Puntuación de sentimientos de 1 a 5 estrellas
- Procesamiento de lenguaje natural con:
 - Detección de palabras positivas/negativas
 - Manejo de modificadores (muy, super, bastante)
 - Reconocimiento de negaciones (no, nunca, jamás)

Visualización y Exportación

- Interfaz web intuitiva con Streamlit
- Tablas interactivas de resultados

- Rankings de productos por sentimiento
- Exportación a CSV
- Métricas y estadísticas en tiempo real

⚡ Optimización de Rendimiento

- Rotación automática de User-Agents
- Pausas inteligentes entre solicitudes
- Manejo robusto de errores
- Estimación de tiempo de procesamiento

🚀 Instalación

Prerrequisitos

- Python 3.7 o superior
- pip (gestor de paquetes de Python)

Instalación de Dependencias

```
bash
```

```
# Clonar el repositorio
```

```
git clone https://github.com/tu-usuario/mercadolibre-scraper.git
```

```
cd mercadolibre-scraper
```

```
# Instalar dependencias
```

```
pip install streamlit pandas urllib3 ssl json statistics datetime collections re csv t
```

Dependencias Principales

```
streamlit>=1.28.0
```

```
pandas>=1.5.0
```

💻 Uso

Ejecutar la Aplicación

```
bash
```

```
streamlit run ml_scraper_v1.py
```




La aplicación se abrirá automáticamente en tu navegador en `http://localhost:8501`

Interfaz de Usuario

1. Panel Lateral de Configuración:

- Término de búsqueda
- Número de páginas (1-5)
- Activar extracción de comentarios
- Número de productos para análisis

2. Pestañas de Resultados:

-  **Datos Extraídos:** Tabla con todos los productos
-  **Análisis de Sentimientos:** Rankings y métricas
-  **Descargas:** Exportación de archivos CSV

Configuración

Parámetros de Scraping

Parámetro	Rango	Descripción
Páginas	1-5	Número de páginas a procesar
Productos	1-250	Cantidad de productos para extraer comentarios
Comentarios	5 por producto	Número fijo de comentarios por producto

Capacidades por Configuración

Páginas	Productos Max	Comentarios Max	Tiempo Estimado
1	50	250	5-8 min
2	100	500	10-15 min
3	150	750	15-20 min
4	200	1,000	20-25 min
5	250	1,250	25-30 min

Análisis de Sentimientos

Algoritmo de Procesamiento

El sistema utiliza un enfoque basado en diccionarios con las siguientes características:

Palabras Clave

- **Positivas:** excelente, bueno, increíble, delicioso, recomendable, etc.
- **Negativas:** malo, horrible, terrible, decepcionante, etc.

Modificadores

- **Intensificadores:** muy, super, bastante, extremadamente

- **Negadores:** no, nunca, jamás, tampoco

Cálculo de Puntuación

python

```
# Fórmula simplificada
sentimiento_base = 3.0
puntuacion_final = sentimiento_base + (promedio_palabras * 2.0)
resultado = max(1.0, min(5.0, puntuacion_final))
```

Métricas Generadas

- **Sentimiento Promedio:** Puntuación calculada (1-5 estrellas)
- **Calificación Original:** Puntuación de MercadoLibre
- **Número de Comentarios:** Cantidad de comentarios analizados
- **Comparación:** Diferencia entre sentimiento calculado y original

Estructura de Datos

Datos de Productos Extraídos

CSV

```
nombre,estrellas,calificaciones,precio,descuento,envio,url,comentario_1,puntuacion_com
"Vino Tinto Malbec 2020",4.5,156,"2500","Sin descuento","Estándar","https://...",
"Excelente vino, muy recomendable","5",...
```

Análisis de Sentimientos

CSV

```
nombre_producto,sentimiento_promedio,calificacion_promedio,precio,estrellas,num_coment
"Vino Tinto Malbec 2020",4.2,4.4,2500,4.5,5,"https://..."
```

Ejemplos

Búsqueda Básica

Término: "vinos malbec"

Páginas: 2

Productos con comentarios: 50

Análisis Completo

Término: "notebooks gaming"

Páginas: 5

Productos con comentarios: 200

Casos de Uso Recomendados



Análisis Rápido (5-10 minutos)

- Término: producto específico
- Páginas: 1-2
- Productos: 20-50
- Uso: Validación rápida de producto



Estudio de Mercado (15-20 minutos)

- Término: categoría amplia
- Páginas: 3-4
- Productos: 100-150
- Uso: Análisis competitivo



Investigación Completa (25-30 minutos)

- Término: nicho específico
- Páginas: 5
- Productos: 200-250
- Uso: Estudio profundo de mercado



Consideraciones

Limitaciones Técnicas

- **Rate Limiting:** Pausas automáticas para evitar bloqueos
- **Estructura Web:** Dependiente de la estructura actual de MercadoLibre
- **Comentarios:** No todos los productos tienen comentarios disponibles

Consideraciones Éticas

- Uso responsable del scraping
- Respeto a los términos de servicio
- No sobrecargar los servidores

Recomendaciones de Uso

- Ejecutar durante horarios de baja actividad
- Monitorear el progreso durante ejecuciones largas
- Guardar resultados frecuentemente

Solución de Problemas

Errores Comunes

Error de Conexión

Error al acceder a URL: [Errno 11001] getaddrinfo failed

Solución: Verificar conexión a internet y reintentar

Sin Productos Encontrados

No se pudieron encontrar productos

Solución:

- Verificar término de búsqueda
- Probar con términos más generales
- Verificar que el producto exista en MercadoLibre Argentina

Timeout de Solicitudes

Timeout error

Solución: Reiniciar la aplicación y reducir número de páginas

Optimización de Rendimiento

Para Conexiones Lentas

- Reducir número de páginas
- Limitar productos con comentarios
- Ejecutar en horarios de menor tráfico

Para Análisis Extensos

- Ejecutar por partes
- Guardar resultados intermedios
- Usar términos específicos

Casos de Uso Avanzados

Análisis Competitivo

python

```
# Comparar productos similares
terminos = ["smartphone samsung", "smartphone iphone", "smartphone xiaomi"]
for termino in terminos:
    # Ejecutar scraping individual
    # Comparar sentimientos promedio
```

Monitoreo de Marca

python

```
# Seguimiento de productos de una marca
termino = "marca_especifica"
# Análisis mensual de sentimientos
# Detección de cambios en percepción
```

Investigación de Mercado

python

```
# Análisis de categorías completas
categorias = ["laptops", "tablets", "smartphones"]
# Identificación de tendencias
# Análisis de precios vs sentimientos
```

Contribución

Cómo Contribuir

1. **Fork** el repositorio
2. **Crea** una rama para tu feature (`git checkout -b feature/AmazingFeature`)
3. **Commit** tus cambios (`git commit -m 'Add some AmazingFeature'`)
4. **Push** a la rama (`git push origin feature/AmazingFeature`)
5. **Abre** un Pull Request

Áreas de Mejora

- ☐ Soporte para otros países de MercadoLibre
- ☐ Análisis de sentimientos más avanzado con ML
- ☐ Interfaz de programación (API)

- ☐ Exportación a otros formatos (Excel, JSON)
- ☐ Visualizaciones gráficas avanzadas
- ☐ Sistema de alertas y notificaciones
- ☐ Base de datos para almacenamiento histórico

Reportar Bugs

Usa las [GitHub Issues](#) para reportar bugs:

- **Describe** el problema claramente
- **Incluye** pasos para reproducir
- **Proporciona** logs de error si están disponibles
- **Especifica** tu entorno (OS, Python version, etc.)



Licencia

Este proyecto está bajo la Licencia MIT - ver el archivo [LICENSE](#) para más detalles.



Agradecimientos

- **Streamlit** por la excelente framework de aplicaciones web
- **Pandas** por el manejo eficiente de datos
- **MercadoLibre** por proporcionar una plataforma rica en datos
- **Comunidad Python** por las librerías y herramientas



Contacto

- **Autor:** Tu Nombre
- **Email:** tu.email@ejemplo.com
- **GitHub:** [@tu-usuario](#)
- **LinkedIn:** [Tu Perfil](#)

<div align="center">

★ ¡Si este proyecto te fue útil, no olvides darle una estrella! ★

[↑ Volver al inicio](#)

</div>