

# Análisis de Sentimientos

César Alexis Gómez Vieyra  
Facultad de ciencias físico matemáticas  
cesar.gomezvi@uanl.edu.mx

**Abstract**—El análisis de sentimientos, también conocido como minería de opiniones, es una técnica de procesamiento del lenguaje natural (PLN) que se utiliza para identificar, extraer y cuantificar las emociones y opiniones expresadas en un texto. Este análisis puede determinar si el sentimiento detrás de un texto es positivo, negativo o neutral, y puede ser aplicado a una variedad de fuentes de datos, como redes sociales, reseñas de productos, comentarios en blogs, y más.

El objetivo principal del análisis de sentimientos es comprender las percepciones y actitudes de las personas hacia un tema, producto, servicio o entidad específica. Esto se logra mediante el uso de algoritmos y modelos de aprendizaje automático que analizan el lenguaje y el contexto en el que se expresan las opiniones.

El análisis de sentimientos es ampliamente utilizado en áreas como el marketing, la atención al cliente, la investigación de mercado y la gestión de la reputación en línea, ya que proporciona información valiosa sobre cómo los consumidores y el público en general perciben una marca o producto.

**Index Terms**—análisis de sentimientos, procesamiento de lenguaje natural, LLM's

## I. INTRODUCCIÓN

En la era digital actual, la cantidad de datos generados por los usuarios en plataformas en línea es inmensa y sigue creciendo exponencialmente. Redes sociales, foros, blogs, reseñas de productos y servicios, entre otros, son fuentes ricas de opiniones y emociones expresadas por millones de personas en todo el mundo. Comprender estas opiniones y emociones es crucial para empresas, investigadores y organizaciones que buscan mejorar sus productos, servicios y estrategias de comunicación, es aquí es donde entra en juego el análisis de sentimientos.

El análisis de sentimientos, también conocido como minería de opiniones, es una técnica de procesamiento del lenguaje natural (PLN) que se utiliza para identificar y extraer información subjetiva de los textos. Su objetivo principal es determinar la actitud del autor del texto hacia un tema específico, clasificando el sentimiento como positivo, negativo o neutral. Esta técnica no solo se limita a la polaridad del sentimiento, sino que también puede identificar emociones más complejas como alegría, tristeza, ira, sorpresa, entre otras.

## II. REVISIÓN DE LA LITERATURA

El campo del análisis de sentimientos y la minería de opiniones ha recibido una atención considerable en la última década debido a su capacidad para extraer y analizar opiniones y sentimientos expresados en textos generados por los usuarios. Un trabajo fundamental en este ámbito es el artículo de

revisión de Bo Pang y Lillian Lee (2008), titulado "Opinion Mining and Sentiment Analysis".

Pang y Lee (2008) definen el análisis de sentimientos y la minería de opiniones como técnicas destinadas a identificar y extraer información subjetiva de textos, con el objetivo de determinar la actitud del autor hacia un tema específico. Estas técnicas son esenciales para comprender las actitudes y emociones de los usuarios hacia productos, servicios y eventos.

## III. METODOLOGÍA

### A. Recopilación de Datos

Los datos utilizados en este estudio fueron obtenidos del conjunto de datos "Amazon Reviews" disponible en Kaggle. Este conjunto de datos contiene 34,686,770 reseñas de Amazon de 6,643,669 usuarios sobre 2,441,053 productos, recopiladas por el Stanford Network Analysis Project (SNAP). El subconjunto utilizado en este análisis contiene 1,800,000 muestras de entrenamiento y 200,000 muestras de prueba para cada polaridad de sentimiento.

El conjunto de datos de reseñas de Amazon consiste en reseñas obtenidas de Amazon. Los datos abarcan un período de 18 años, incluyendo aproximadamente 35 millones de reseñas hasta marzo de 2013. Las reseñas incluyen información sobre el producto y el usuario, calificaciones y una reseña en texto plano. Para más información, se puede consultar el siguiente artículo: J. McAuley y J. Leskovec. "Hidden factors and hidden topics: understanding rating dimensions with review text." RecSys, 2013.

### B. Preprocesamiento de Datos

- **Carga de Datos:** Los datos fueron cargados en un DataFrame de pandas, una estructura de datos que permite una manipulación eficiente y análisis de datos tabulares. Se seleccionaron las columnas relevantes que contenían las reseñas de los productos, en este caso se unieron las columnas de la reseña corta y larga

	tokens
0	Despite the fact that I have only played a sma...
1	I bought this charger in Jul 2003 and it worke...
2	Check out Maha Energy's website. Their Powerex...
3	Reviewed quite a bit of the combo players and ...
4	I also began having the incorrect disc problem...

TABLE I  
PRIMEROS 5 REGISTROS

- **Muestreo:** Dado el tamaño considerable del conjunto de datos original, se seleccionó una muestra aleatoria de

10,000 reseñas para el análisis. Este muestreo permite manejar los datos de manera más eficiente y reducir el tiempo de procesamiento sin sacrificar la representatividad de los datos.

- **Tokenización y Limpieza:** Se definió un proceso de tokenización para dividir el texto en palabras individuales (tokens). Durante este proceso, se eliminaron los signos de puntuación y las palabras vacías (stopwords) que no aportan valor semántico significativo al análisis. La limpieza del texto también incluyó la conversión de todas las palabras a minúsculas para asegurar la uniformidad.
- **Explosión de Tokens:** Para facilitar el análisis de sentimientos a nivel de palabra, se transformó el DataFrame de manera que cada token ocupe una fila separada. Esto permite un análisis más granular de los términos individuales en las reseñas.

	tokens	token
0	I bought the Poulan Pro 2 Summer's ago, and it...	bought
0	I bought the Poulan Pro 2 Summer's ago, and it...	poulan
0	I bought the Poulan Pro 2 Summer's ago, and it...	pro
0	I bought the Poulan Pro 2 Summer's ago, and it...	2
0	I bought the Poulan Pro 2 Summer's ago, and it...	summers

TABLE II

TABLA DE TOKEN POR FRASE

### C. Modelos de Análisis de Sentimientos

- **Carga del Léxico de Sentimientos:** Se utilizó el léxico AFINN, un recurso ampliamente utilizado en el análisis de sentimientos que asigna una puntuación de sentimiento a cada término. Este léxico contiene palabras con anotaciones de sentimiento que van desde -5 (muy negativo) hasta +5 (muy positivo).

ID	termino	sentimiento
3094	tyrannic	-3
2639	sabotage	-2
241	authority	1
2929	succeeding	3
2506	refine	1
175	apologizes	-1
225	atrocitiy	-3
1313	flagship	2
1276	favourably	2
585	conflicting	-2

TABLE III

MUESTRA ALEATORIA DE TÉRMINOS Y SENTIMIENTOS

- **Cálculo de Sentimientos:** Se realizó una fusión entre los tokens obtenidos del preprocesamiento y el léxico de sentimientos. Posteriormente, se calculó la media de las puntuaciones de sentimiento para cada reseña. Este enfoque permite obtener una medida agregada del sentimiento expresado en cada reseña, proporcionando una visión general de la polaridad del sentimiento (positivo, negativo o neutral).

ID	tokens	sentimiento
212	AS A FUTURE MEDICAL PROFESSIONAL I THINK THAT ...	0.857143
1774	I am a native Portuguese speaker and I bought ...	1.000000
8225	This is way the best sammy hagar era van halen...	2.583333
1449	Great book. Explains tech terms. Ties in evolu...	2.000000
1041	Don't be seduced by Pinkermania; he's trapped ...	-0.333333
28	"The Van" is one of my favorite books, but I r...	0.000000
5414	My flock of lovebirds love Harrison's products...	1.500000
6968	The toaster has an adjustment for dark to ligh...	-0.333333
4934	It's not the worst Beatles covers ever, but th...	-0.285714
3517	I purchased the flag as a replacement for one ...	2.333333

TABLE IV

TABLA APLICANDO EL PROMEDIO DE LOS SENTIMIENTO DE CADA TOKEN

### D. Visualización de Resultados

Se generó un gráfico de barras para mostrar la distribución de las puntuaciones de sentimiento en las reseñas. Este gráfico proporciona una representación visual de la frecuencia de diferentes niveles de sentimiento, permitiendo una comprensión rápida de la tendencia general de las opiniones.

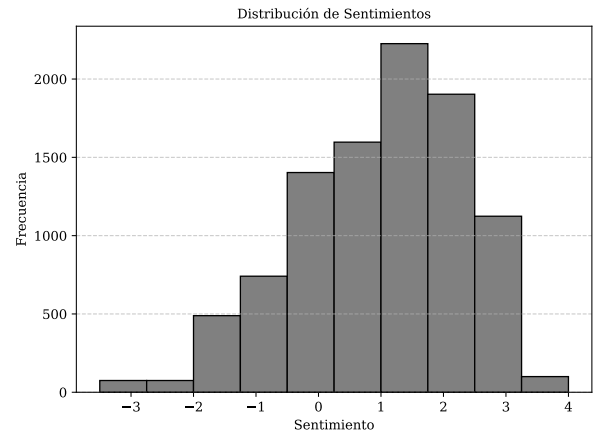


Fig. 1. Distribución de Sentimientos por reseña

## IV. RESULTADOS

El análisis de los sentimientos de las reseñas en Amazon muestra que la media de los sentimientos es aproximadamente 0.93, lo que sugiere que, en promedio, las reseñas tienden a ser ligeramente positivas. La mediana de 1.0 indica que al menos la mitad de las reseñas tienen un sentimiento positivo, mientras que la moda de 2.0 revela que el valor más frecuente entre las reseñas es muy positivo. La desviación estándar de 1.34 sugiere una variabilidad moderada en los sentimientos, indicando que las opiniones de los usuarios varían considerablemente. El boxplot muestra que los sentimientos se distribuyen desde aproximadamente -3 hasta 4, con algunos valores atípicos negativos alrededor de -3, lo que representa reseñas con sentimientos extremadamente negativos. Sin embargo, estos valores atípicos son pocos en comparación con el total de reseñas. La mayoría de las reseñas se agrupan en el rango positivo, como lo indica el rango intercuartílico que se extiende desde 0 hasta 2. En resumen, aunque hay

una diversidad considerable en las opiniones de los usuarios, con algunas reseñas muy negativas, la tendencia general es hacia sentimientos positivos, lo que sugiere una percepción mayormente favorable de los productos reseñados en Amazon.

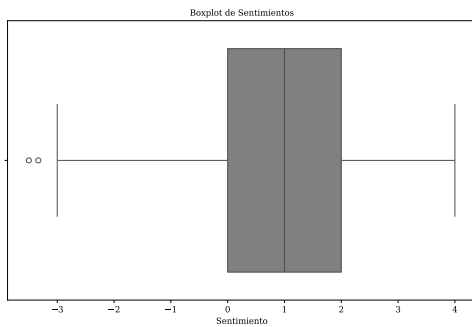


Fig. 2. Boxplot de Sentimientos por reseña

## V. CONCLUSIONES

El análisis de los sentimientos de las reseñas en Amazon revela una tendencia general hacia opiniones positivas. Con una media de 0.93 y una mediana de 1.0, es evidente que la mayoría de los usuarios tienen una percepción favorable de los productos que han evaluado. La moda de 2.0, que representa el valor más frecuente, refuerza esta tendencia positiva. Sin embargo, la desviación estándar de 1.34 indica que existe una variabilidad considerable en las opiniones, lo que sugiere que no todos los usuarios comparten la misma experiencia. El boxplot muestra que, aunque la mayoría de las reseñas se encuentran en el rango positivo, hay algunos valores atípicos negativos que representan experiencias extremadamente desfavorables. Estos valores atípicos, aunque menos frecuentes, son importantes ya que destacan la diversidad de opiniones y la existencia de experiencias negativas significativas.

El análisis de sentimientos, una herramienta poderosa para comprender las emociones y percepciones de los usuarios, ha permitido identificar estas tendencias y variaciones en las reseñas. En conjunto, estos hallazgos sugieren que, aunque la percepción general de los productos en Amazon es positiva, las experiencias de los usuarios varían ampliamente. Esto subraya la importancia de considerar tanto las opiniones positivas como las negativas al evaluar la calidad y el desempeño de los productos. El análisis de sentimientos proporciona una visión más profunda y matizada de las opiniones de los usuarios, lo que es crucial para empresas y consumidores que buscan tomar decisiones informadas.

## REFERENCES

- [1] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- [2] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135. <https://doi.org/10.1561/15000000011>
- [3] Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 168-177. <https://doi.org/10.1145/1014052.1014073>