

# Tarea 2 - Análisis de Sentimiento y Preprocesamiento de Datos de Productos

Nelson Alfonso Beas Ham

Matrícula: 1942687

Universidad Autónoma de Nuevo León

Facultad de Ciencias Físico Matemáticas

29 de enero de 2025

## 1. Introducción

En este reporte se realiza un análisis de sentimiento sobre una lista de nombres de productos utilizando el modelo **TextBlob**. El objetivo principal es aplicar técnicas de procesamiento de texto, como la limpieza de datos, tokenización, eliminación de palabras vacías y análisis de sentimientos. Además, se presentan visualizaciones de los resultados obtenidos para mostrar la distribución de los sentimientos en los productos analizados.

## 2. Metodología

El análisis se llevó a cabo siguiendo los siguientes pasos:

1. **Carga de Datos:** Se cargó el conjunto de datos desde un archivo CSV que contiene los nombres de los productos.
2. **Preprocesamiento:** Se limpió el texto de los nombres de productos, eliminando caracteres especiales y convirtiéndolos a minúsculas.
3. **Análisis de Sentimiento:** Se utilizó el modelo de **TextBlob** para determinar el sentimiento de cada nombre de producto, clasificándolos en categorías como *Positivo* o *Negativo*.
4. **Visualización:** Se generaron gráficos para visualizar los resultados del análisis de sentimiento y la distribución de los sentimientos en los nombres de productos.

### 3. Preprocesamiento

El preprocesamiento de los datos consistió en limpiar los nombres de los productos para hacerlos aptos para el análisis de sentimiento. Las operaciones realizadas fueron las siguientes:

- Eliminación de caracteres especiales como puntuaciones.
- Conversión de todos los caracteres a minúsculas.
- Reemplazo de caracteres no alfabéticos.
- Creación de una nueva columna que contenía los nombres de productos procesados.

### 4. Análisis de Sentimiento

El análisis de sentimiento fue realizado con `TextBlob`, una librería de Python que permite obtener el puntaje de polaridad de un texto. Los puntajes de polaridad van desde -1 (sentimiento negativo) hasta 1 (sentimiento positivo). Para clasificar los sentimientos de los productos, se utilizaron las siguientes categorías:

- **Positivo:** Si el puntaje de polaridad es mayor que 0.
- **Negativo:** Si el puntaje de polaridad es menor o igual a 0.

La función `simple_sentiment_analysis` fue utilizada para aplicar este análisis a cada nombre de producto.

Listing 1: Función de Análisis de Sentimiento con TextBlob

```
from textblob import TextBlob

# Función para análisis de sentimiento con TextBlob
def simple_sentiment_analysis(text):
    analysis = TextBlob(text)
    # Determina el sentimiento: positivo si la polaridad es mayor que 0
    return 'POSITIVE' if analysis.sentiment.polarity > 0 else 'NEGATIVE'

# Aplicar análisis de sentimiento a todo el dataframe
df['simple_sentiment'] = df['product_name'].apply(simple_sentiment_analysis)

# Mostrar las primeras filas con el análisis de sentimiento
print(df[['product_name', 'simple_sentiment']].head())
```

# 5. Visualización

Para analizar los resultados obtenidos del análisis de sentimiento, se generaron visualizaciones que muestran la distribución de los sentimientos en los nombres de productos.

## 5.1. Distribución de Sentimientos

A continuación, se presenta el gráfico que muestra la distribución de los sentimientos en los nombres de los productos. Este gráfico representa cuántos productos fueron clasificados como *Positivos* y cuántos como *Negativos*.

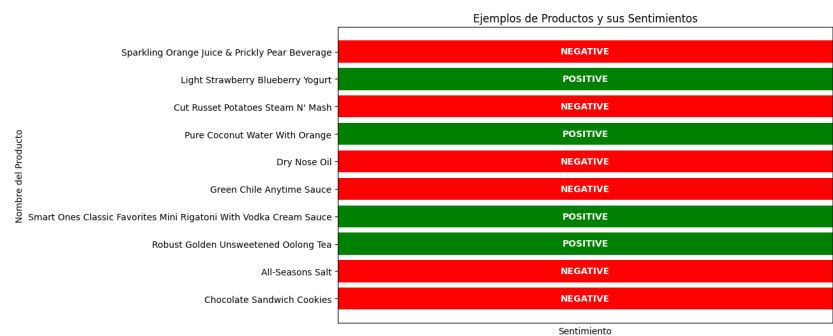


Figura 1: Ejemplos de sentimientos asignados

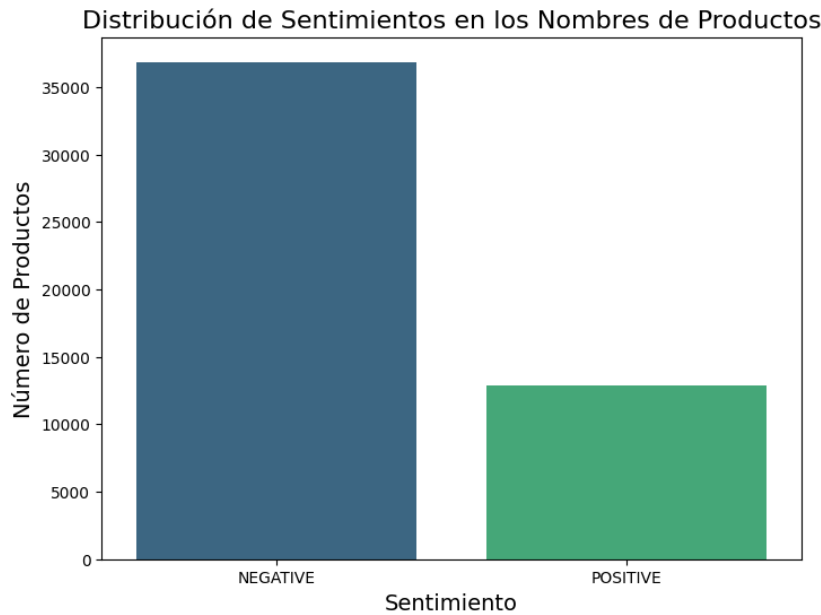


Figura 2: Distribución de Sentimientos con Gráfico de Barras

## 5.2. Ejemplos de Productos y Sentimientos

Se muestran ejemplos de productos con su respectivo sentimiento asignado. Este gráfico ilustra cómo los nombres de los productos son percibidos emocionalmente, lo que puede ser útil para evaluar su efectividad en términos de marketing.

## 6. Conclusiones

El análisis de los nombres de productos permitió identificar la polaridad de cada uno, clasificándolos en sentimientos positivos y negativos. Los resultados de este análisis pueden ser útiles para el desarrollo de estrategias de marketing y branding, ya que proporcionan una visión general de cómo los nombres de productos son percibidos emocionalmente por los consumidores. Además, las visualizaciones generadas permiten una comprensión más clara de la distribución de sentimientos en los datos.

En futuros trabajos, se podría explorar la posibilidad de utilizar modelos más avanzados de análisis de sentimientos que puedan considerar matices adicionales en los textos y mejorar la clasificación de los productos.