Optimización de campañas publicitarias a través del análisis de sentimientos: Un enfoque basado en algoritmos de clasificación

Jesús David Arévalo Montilla

2024

Abstract

En la era digital, la eficacia de las campañas publicitarias es crucial para el éxito de las marcas en un mercado altamente competitivo. Este artículo explora cómo el análisis de sentimientos, implementado mediante algoritmos de clasificación, puede optimizar las estrategias publicitarias. Utilizando técnicas avanzadas de procesamiento de lenguaje natural y machine learning, se analizan grandes volúmenes de datos provenientes de la red social de twitter, donde reseñas de productos y comentarios de usuarios pueden ayudar a identificar patrones y tendencias en las percepciones de los consumidores. Los resultados de un estudio así pueden demostrar que la incorporación del análisis de sentimientos permite una segmentación más precisa del mercado, una personalización de los mensajes publicitarios y una mejora en la toma de decisiones estratégicas. Este estudio destaca el potencial del análisis de sentimientos como una herramienta poderosa para mejorar la efectividad y el retorno de inversión de las campañas publicitarias, proporcionando a los profesionales de marketing un enfoque basado en datos para abordar las necesidades y preferencias de los consumidores.

Keywords: Análisis de Sentimientos, Algoritmos de Clasificación, Campañas Publicitarias, Optimización de Publicidad, Machine Learning, Estrategias de Marketing, Opinión del Consumidor.

1 Introducción

En un mercado cada vez más competitivo, la capacidad de captar y mantener la atención de los consumidores es crucial para el éxito de las campañas publicitarias. El análisis de sentimientos, basado en algoritmos de clasificación, ofrece una herramienta poderosa para mejorar la eficacia de estas campañas. Las opiniones y reacciones de los consumidores contienen información clave para determinar las acciones que debe tomar un equipo de marketing. En la actualidad, marcar tendencia es primordial, y esto solo se logra generando un impacto significativo en el público objetivo.

2 Datos

2.1 Fuente de Datos

En este paper, se utilizaron datos provenientes de redes sociales, específicamente de Twitter, donde se analizaron los posts realizados por diversos usuarios en el año 2009. El dataset Sentiment140 encontrado en Kaggle, contiene 1,600,000 tweets extraídos mediante la API de Twitter, fue empleado para este análisis. Estos tweets han sido anotados con polaridades de sentimiento (0 = negativo, 4 = positivo) y pueden ser utilizados para detectar sentimientos. El dataset incluye seis campos: la polaridad del tweet, el identificador único del tweet, la fecha y hora de publicación, la consulta utilizada para extraer el tweet, el nombre de usuario que publicó el tweet y el contenido del tweet. Esta información permite estudiar cómo los usuarios expresan sus emociones en redes sociales y cómo estas expresiones pueden ser clasificadas automáticamente mediante algoritmos de clasificación y procesamiento de lenguaje natural.

2.2 Análisis Exploratorio de Datos (EDA)

Primeramente, el dataset se modificó en dos partes: una para mejorar el procesamiento durante el entrenamiento de los modelos y la otra para mejorar la comprensión de las categorías de tipo de comentario a las cuales pertenece cada post. Las dimensiones del dataset trabajadas componen el 10% del dataset inicial, es decir, nuestra muestra fue de 160,000 tweets, suficiente para realizar el ejercicio de investigación sin inconvenientes con el poder de cómputo de la máquina utilizada para la evaluación de los modelos. Otro cambio realizado fue que, en vez de usar el número 4 para clasificar los comentarios positivos, se hace uso del uno, logrando así una clasificación más binaria. La categoría de comentarios neutros fue omitida.

2.2.1 Polaridad del Sentimiento (target)

- **Cuenta**: Hay 160,000 registros en total.
- **Media**: La media es de 0.501175, lo que sugiere que hay un balance casi perfecto entre tweets positivos (valor 1) y negativos (valor 0).
- **Desviación Estándar**: La desviación estándar es de 0.5, indicando una dispersión igual entre los valores 0 y 1.
- **Mínimo y Máximo**: Los valores oscilan entre 0 (negativo) y 1 (positivo).
- **Percentiles**: El 25% de los tweets tienen una polaridad de 0, el 50% (mediana) tienen una polaridad de 1, y el 75% también tienen una polaridad de 1.

2.2.2 Identificadores de los Tweets (ids)

- **Cuenta**: Hay 160,000 registros únicos.
- **Media**: La media de los identificadores es aproximadamente 1,999,154,000.
- **Desviación Estándar**: La desviación estándar es de aproximadamente 193,504,700, lo que indica una considerable dispersión en los identificadores de los tweets.

- **Mínimo y Máximo**: Los valores de identificadores varían entre 1,467,814,000 y 2,329,205,000.
- **Percentiles**: El 25% de los tweets tienen identificadores menores a 1,957,035,000, el 50% tienen identificadores menores a 2,002,126,000, y el 75% tienen identificadores menores a 2,177,168,000.

2.2.3 Fechas de los Tweets (date)

- **Cuenta**: Hay 160,000 fechas registradas.
- **Media**: La fecha promedio es alrededor del 31 de mayo de 2009.
- **Rango de Fechas**: Las fechas van desde el 6 de abril de 2009 hasta el 25 de junio de 2009.
- **Percentiles**: El 25% de los tweets fueron publicados antes del 28 de mayo de 2009, el 50% fueron publicados antes del 2 de junio de 2009, y el 75% fueron publicados antes del 15 de junio de 2009.

En resumen, el dataset se modificó para mejorar tanto el procesamiento como la comprensión de las categorías de comentarios, utilizando una muestra del 10% del dataset original, es decir, 160,000 tweets. Se ajustó la clasificación de los sentimientos, usando 0 para negativo y 1 para positivo, omitiendo la categoría neutra para simplificar el análisis binario. La distribución de la polaridad del sentimiento mostró un balance casi perfecto entre tweets positivos y negativos, con una media de 0.501175 y una desviación estándar de 0.5. Los identificadores de los tweets presentaron una considerable dispersión, con valores que oscilan entre 1,467,814,000 y 2,329,205,000, y una media de aproximadamente 1,999,154,000. Las fechas de los tweets abarcan desde el 6 de abril de 2009 hasta el 25 de junio de 2009, con una fecha promedio alrededor del 31 de mayo de 2009. Estos ajustes y análisis permiten un procesamiento más eficiente y una interpretación más clara de los datos para la evaluación de los modelos.

2.3 Análisis Gráfico:

Figure 1: target 0: Comentarios negativos, target 1: Comentarios positivos

Descripción: Este diagrama de barras muestra la distribución de tweets clasificados como positivos (1) y negativos (0).

Interpretación: La distribución es casi balanceada, con un número ligeramente mayor de tweets negativos comparados con los positivos. Esta distribución equilibrada es ideal para entrenar modelos de clasificación de sentimientos, ya que proporciona una cantidad similar de ejemplos para cada categoría.

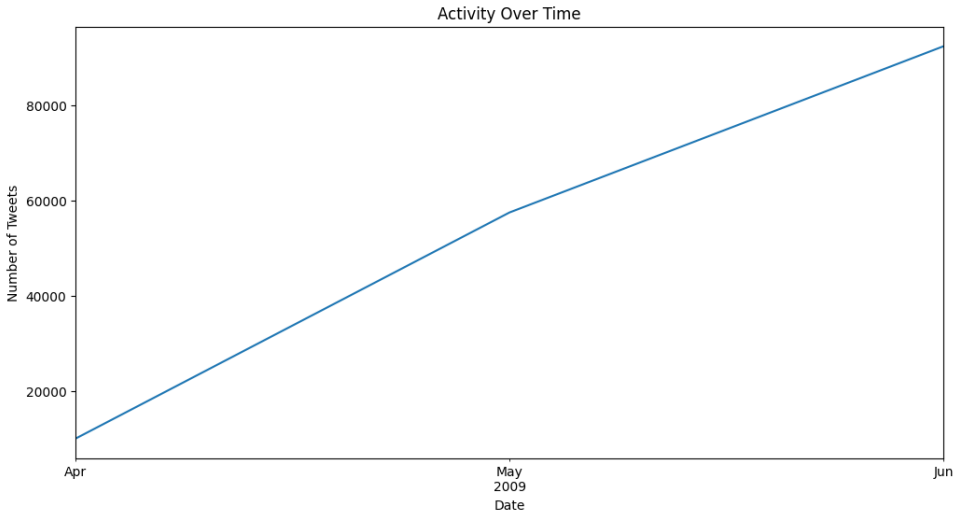

Figure 2: Actividad en el tiempo

Descripción: Este gráfico de líneas muestra la actividad total de tweets a lo largo del tiempo en 2009.

Interpretación: El gráfico muestra un incremento constante en la actividad de tweets desde abril hasta junio de 2009. Este aumento puede reflejar un crecimiento en el uso de

Twitter o en la cantidad de datos recopilados durante este período. Esta información es útil para entender el comportamiento general de los usuarios y la evolución de la actividad en la plataforma.

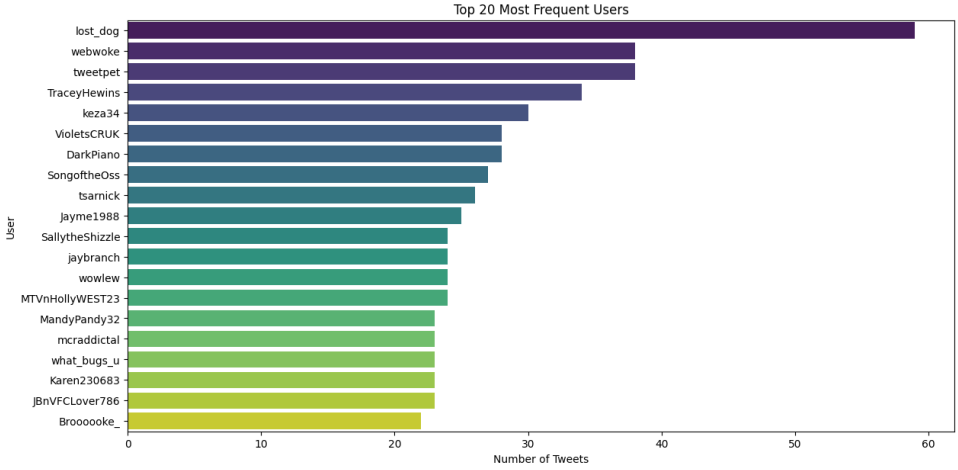

Figure 3: Usuarios frecuentes

Descripción: Este gráfico de barras muestra los 20 usuarios más frecuentes en el dataset. Cada barra representa el número de tweets publicados por cada usuario.

Interpretación: El usuario *lost_dog* es el más frecuente, seguido por otros como "webwoke" y "tweetpet". Este análisis puede ayudar a identificar usuarios clave y potenciales influencers dentro de la muestra de datos, proporcionando información valiosa para estrategias de marketing dirigidas.

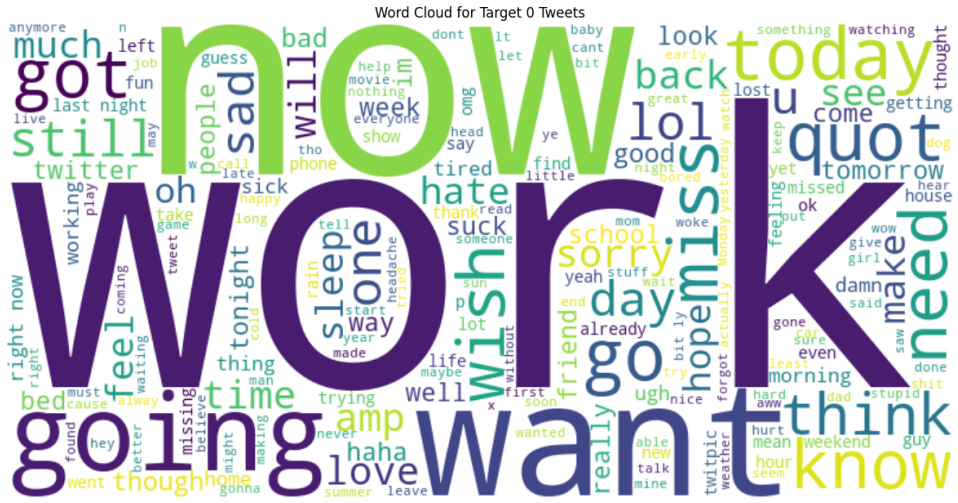

Figure 4: Nube Palabras negativas

Descripción: La imagen muestra una nube de palabras que representa la frecuencia de las palabras en tweets con sentimiento negativo (target 0). Las palabras más grandes y destacadas, como "now", "work", "want", "sad", y "going", indican términos que son comúnmente utilizados en tweets negativos. Este análisis visual ayuda a identificar los temas y emociones más prevalentes en los tweets negativos, como frustración, trabajo, deseo y tristeza.

Interpretación: Las palabras más frecuentes reflejan preocupaciones diarias y estados emocionales negativos. Esto puede ser útil para entender qué términos están más asociados con sentimientos negativos y puede guiar a las estrategias de análisis de sentimiento y de marketing para abordar estos temas.

Figure 5: Nube Palabras Positivas

Descripción: La imagen muestra una nube de palabras para tweets con sentimiento positivo (target 1). Las palabras más prominentes, como "love", "thank", "today", "happy", y "good", indican términos frecuentemente utilizados en tweets positivos.

Interpretación: Las palabras más frecuentes reflejan emociones positivas y gratitud, como amor, agradecimiento y felicidad. Este análisis puede ser útil para identificar temas y términos asociados con sentimientos positivos, ayudando a mejorar las estrategias de marketing y la comunicación positiva.

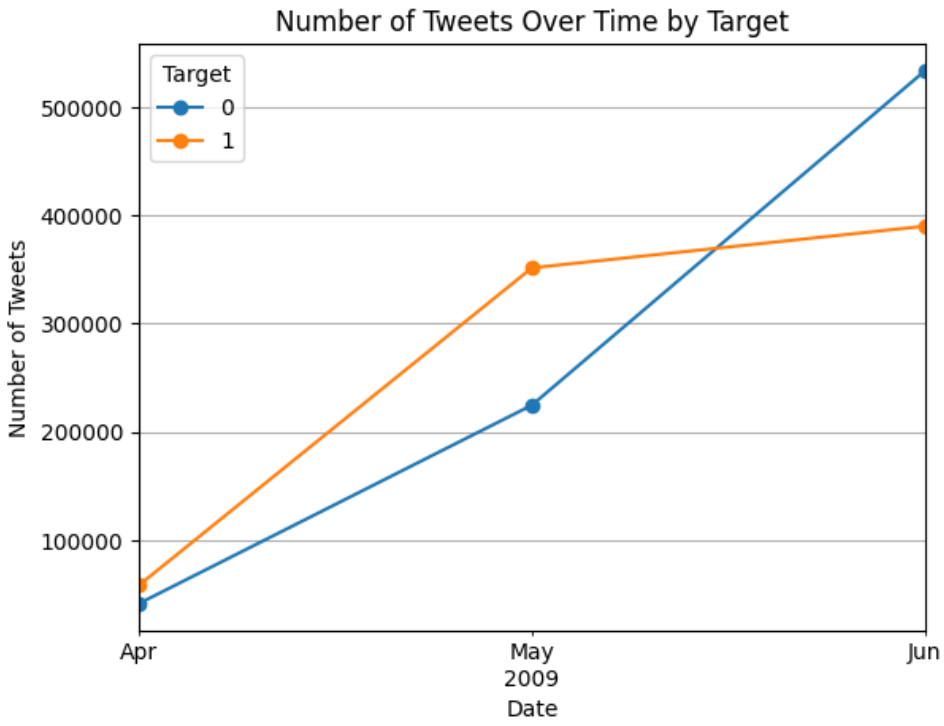

Figure 6: Posts Positivos vs Negativos en el tiempo

Descripción: Este gráfico de líneas muestra el número de tweets clasificados como positivos y negativos a lo largo del tiempo en 2009. La línea azul representa los tweets negativos, mientras que la línea naranja representa los tweets positivos.

Interpretación: Se observa un incremento constante en el número de tweets para ambas categorías a lo largo del tiempo, lo que puede indicar un aumento en la actividad en Twitter durante este período. Los puntos en el gráfico indican momentos específicos donde los tweets positivos superan a los negativos y viceversa. Este tipo de análisis puede ayudar a entender patrones temporales en la actividad de los usuarios y en el sentimiento expresado en los tweets.

3 Modelado de los datos

3.1 Implementación de Modelos y evaluación de Modelos

Los modelos fueron evaluados utilizando métricas como precisión, exactitud, recall y F1-score. Los modelos utilizados para este proyecto fueron Regresión Ridge, Random Forest y KNN. Estos mismos se destacaron por presentar valores significativos en las métricas con las que fueron evaluados.

Table 1: Evaluación del Rendimiento de los Modelos

Modelos	Negativos			Positivos		
	Precisión	Recall	F1-Score	Precisión	Recall	F1-Score
Ridge	0.784	0.742	0.763	0.757	0.797	0.776
2-KNN	0.588	0.765	0.665	0.671	0.471	0.553
Random Forest	0.766	0.758	0.762	0.762	0.770	0.766

3.2 Modelo propuesto

El modelo que se propuso para el desarrollo de este trabajo académico fue XGBoost, no obstante, como podemos ver más adelante en la comparativa realizada con los modelos anteriores, presenta cierta disminución en los valores de las métricas. Alterno a este, se evaluó también KNN, haciendo uso del Square Root of N rule, que ampliaba considerablemente el rango donde se podía encontrar el número de k-vecinos más óptimo para el modelo. La forma en que se desarrolló este método fue que la raíz cuadrada de la muestra total (160,000) compuso el límite superior de búsqueda para este parámetro, ello amplió el orden de complejidad del algoritmo lo cual dejó esta opción de modelo fuera de consideración.

3.3 Selección del Modelo

Dado lo anterior, el modelo que fue seleccionado para estudio fue Ridge, teniendo scores altos tanto a la hora de clasificar los comentarios negativos como los positivos. Este mismo no solo destacó por las métricas evaluadas sino que en términos de costo computacional fue uno de los más eficientes teniendo en cuenta el volumen de los datos con los que fue entrenado. La tabla comparativa del performance incluyendo XGBoost queda de la siguiente manera:

Table 2: Evaluación del Rendimiento de los Modelos

Modelos	Negativos			Positivos		
	Precisión	Recall	F1-Score	Precisión	Recall	F1-Score
XGBoost	0.786	0.622	0.695	0.695	0.834	0.756
Ridge	0.784	0.742	0.763	0.757	0.797	0.776
2-KNN	0.588	0.765	0.665	0.671	0.471	0.553
Random Forest	0.766	0.758	0.762	0.762	0.770	0.766

Interpretación del Gráfico Comparativo de Modelos

El gráfico comparativo de modelos muestra las puntuaciones de varias métricas de evaluación (Accuracy, Precision, Recall, F1-score) para los modelos XGBoost, Ridge, KNN, y Random Forest. Cada línea en el gráfico representa una métrica diferente, permitiendo una visualización clara de las diferencias de rendimiento entre los modelos.

Observaciones Clave

1. **Accuracy**:
    - Los modelos Ridge y Random Forest tienen las puntuaciones de accuracy más altas y similares, seguidos por XGBoost, mientras que KNN tiene la puntuación más baja.
2. **Precision (0)**:
    - XGBoost tiene la precisión más alta para la clase negativa, seguido de cerca por Ridge y Random Forest. KNN tiene la precisión más baja en esta métrica.
3. **Recall (0)**:

- KNN destaca con el recall más alto para la clase negativa, lo que indica su capacidad para identificar correctamente la mayoría de los comentarios negativos. Sin embargo, otros modelos tienen un recall más equilibrado y consistente.

4. **F1-score (0)**:

   - Ridge y Random Forest tienen puntuaciones de F1-score muy similares y altas para la clase negativa, lo que indica un buen equilibrio entre precisión y recall. XGBoost sigue con una puntuación moderada, mientras que KNN tiene la más baja.

5. **Precision (1)** y **Recall (1)**:

   - Los modelos muestran un rendimiento relativamente equilibrado en estas métricas para la clase positiva, con XGBoost teniendo un recall notablemente alto para la clase positiva.

6. **F1-score (1)**:

   - Las puntuaciones son bastante consistentes entre Ridge, Random Forest y XGBoost, con KNN nuevamente siendo el más bajo.

Recomendación

El gráfico refuerza la conclusión de que **Random Forest** y **Ridge** son las mejores opciones, especialmente en términos de F1-score y recall para la clase negativa. Aunque KNN tiene un buen recall para la clase negativa, su baja precisión y F1-score lo hacen menos adecuado. **Random Forest** ofrece un equilibrio excelente en todas las métricas, lo que lo convierte en la opción preferida para clasificar correctamente los comentarios negativos.

Interpretación de la Matriz de Comparación de Modelos

La matriz de comparación de modelos proporciona una visualización detallada de las métricas de evaluación (Accuracy, Precision, Recall, F1-score) para los modelos XGBoost, Ridge, KNN y Random Forest. Cada celda en la matriz representa el valor de una métrica específica para un modelo particular, con colores que indican la magnitud de los valores.

Observaciones Clave

1. **Accuracy**:

   - Ridge y Random Forest tienen las puntuaciones más altas, seguidos de XGBoost y KNN.

2. **Precision (0)**:

   - XGBoost tiene la precisión más alta, seguido por Ridge y Random Forest. KNN tiene la precisión más baja.

3. **Recall (0)**:

- KNN tiene el recall más alto para la clase negativa, lo que indica que es muy efectivo para identificar los comentarios negativos. Sin embargo, su precisión es baja.

4. **F1-score (0)**:
    - Ridge y Random Forest tienen los F1-scores más altos, seguidos por XGBoost y KNN.
5. **Precision (1)** y **Recall (1)**:
    - XGBoost tiene el recall más alto para la clase positiva, mientras que Ridge y Random Forest tienen un rendimiento más equilibrado en ambas métricas.
6. **F1-score (1)**:
    - Los F1-scores son más altos para Ridge, Random Forest y XGBoost, con KNN siendo el más bajo.

Conclusión General

Ambas visualizaciones confirman que **Random Forest** y **Ridge** son los modelos más robustos y consistentes, especialmente para la clasificación de comentarios negativos. **Random Forest** es particularmente destacado por su equilibrio entre precisión y recall, proporcionando un rendimiento sólido en todas las métricas. Por lo tanto, se recomienda seleccionar el modelo Random Forest para clasificar correctamente los comentarios negativos, con Ridge como una alternativa igualmente fuerte. KNN, aunque tiene un buen recall para la clase negativa, su baja precisión y F1-score lo hacen menos confiable para esta tarea. XGBoost, aunque competitivo, no supera a Random Forest y Ridge en términos de equilibrio y consistencia en las métricas evaluadas.

3.4 Grafico Matrices de confusión:

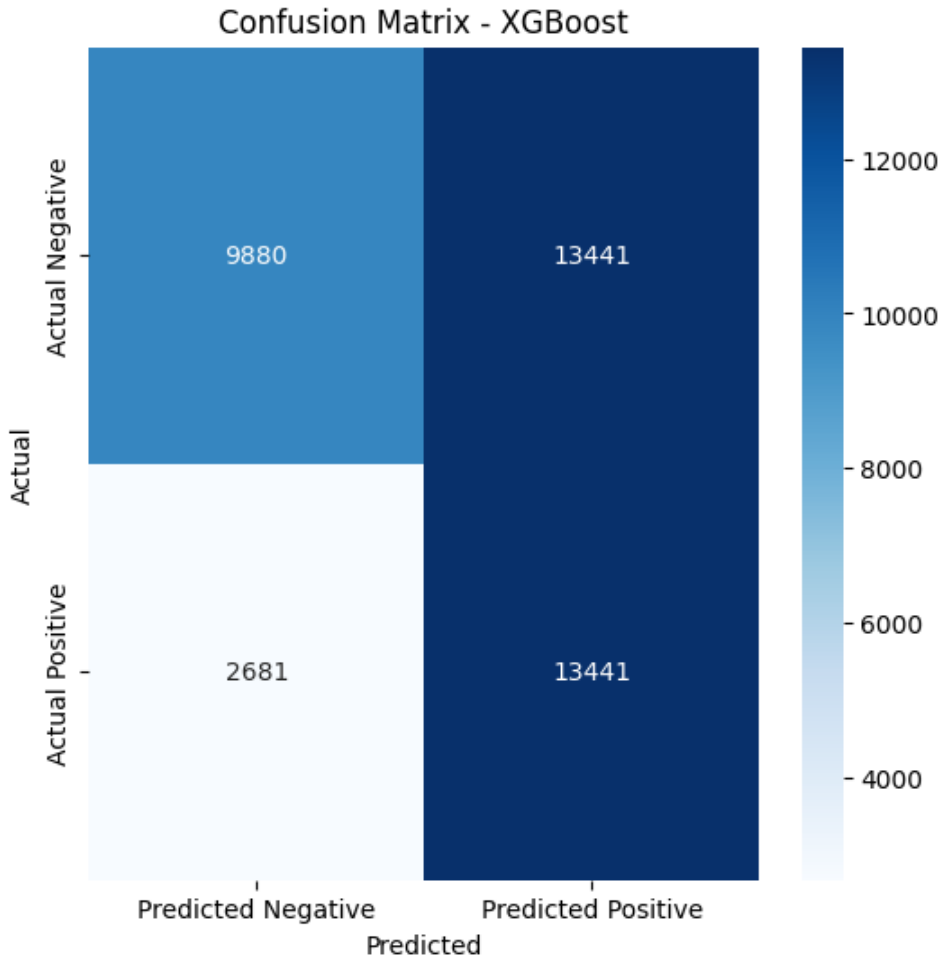

Figure 7: XgBoost

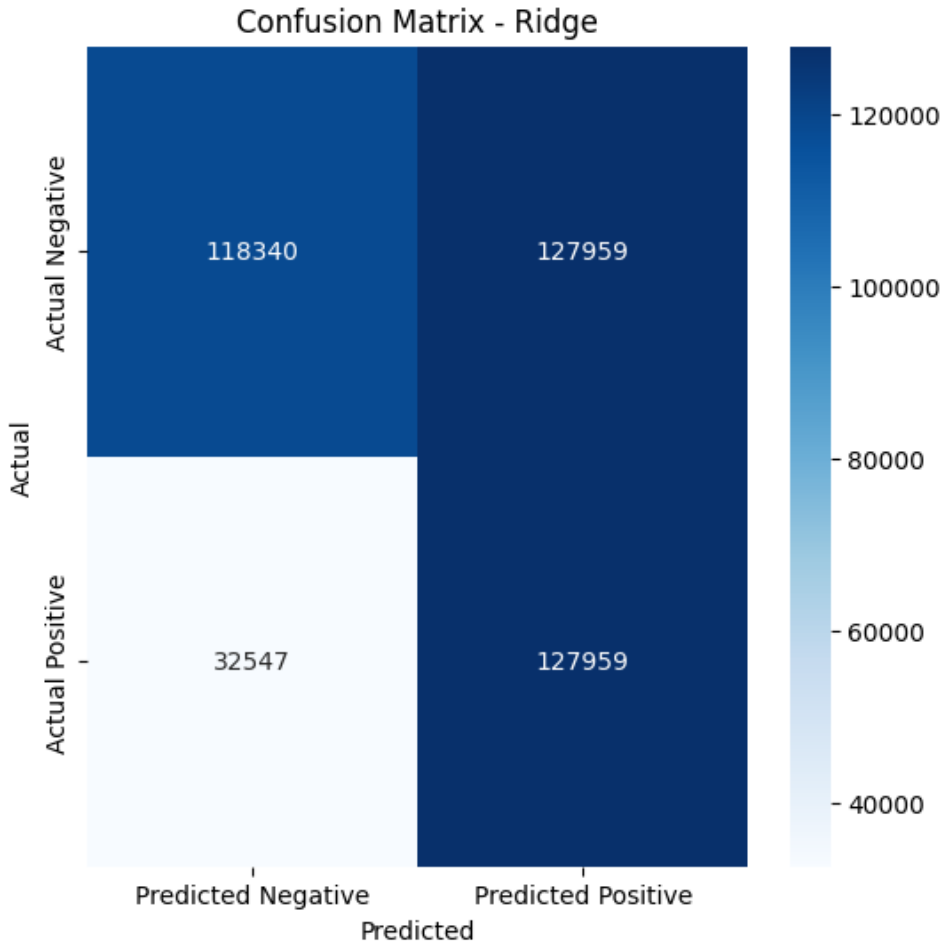

Figure 8: Ridge

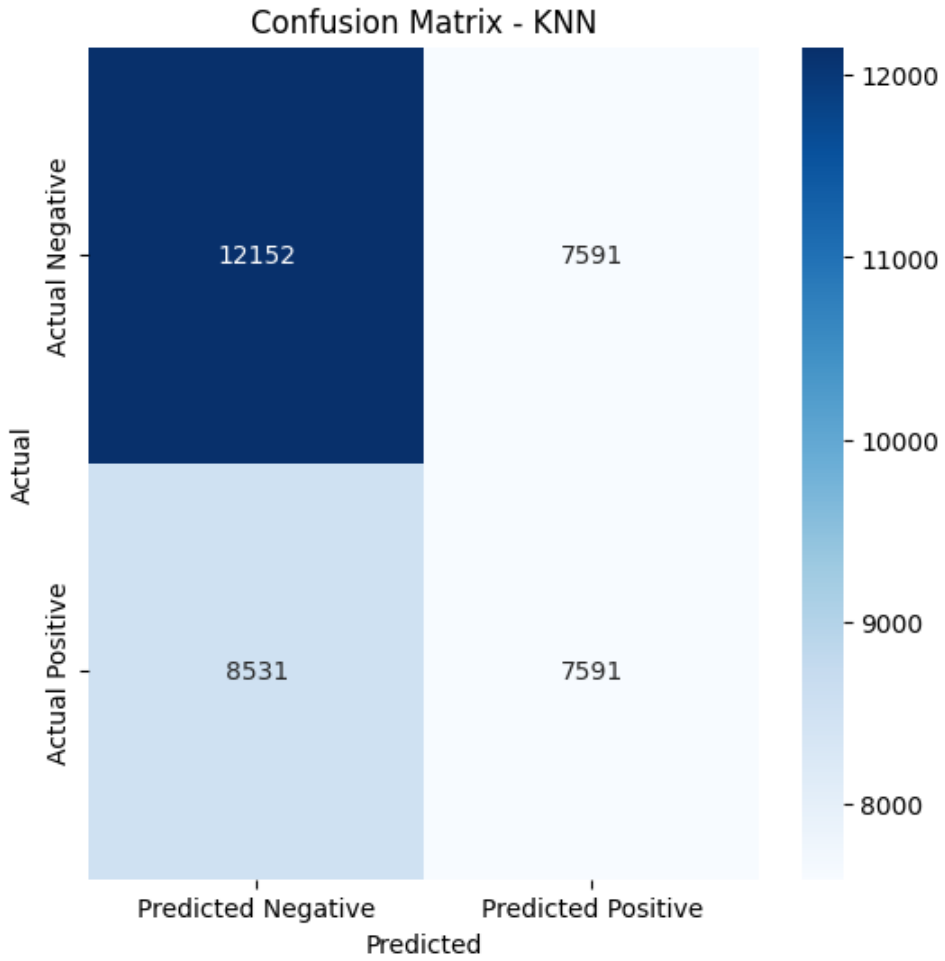
Confusion Matrix - KNN
Actual Negative
Actual Positive
Actual
12152
7591
8531
7591
Predicted Negative
Predicted Positive
Predicted
12000
11000
10000
9000
8000

Figure 9: Knn

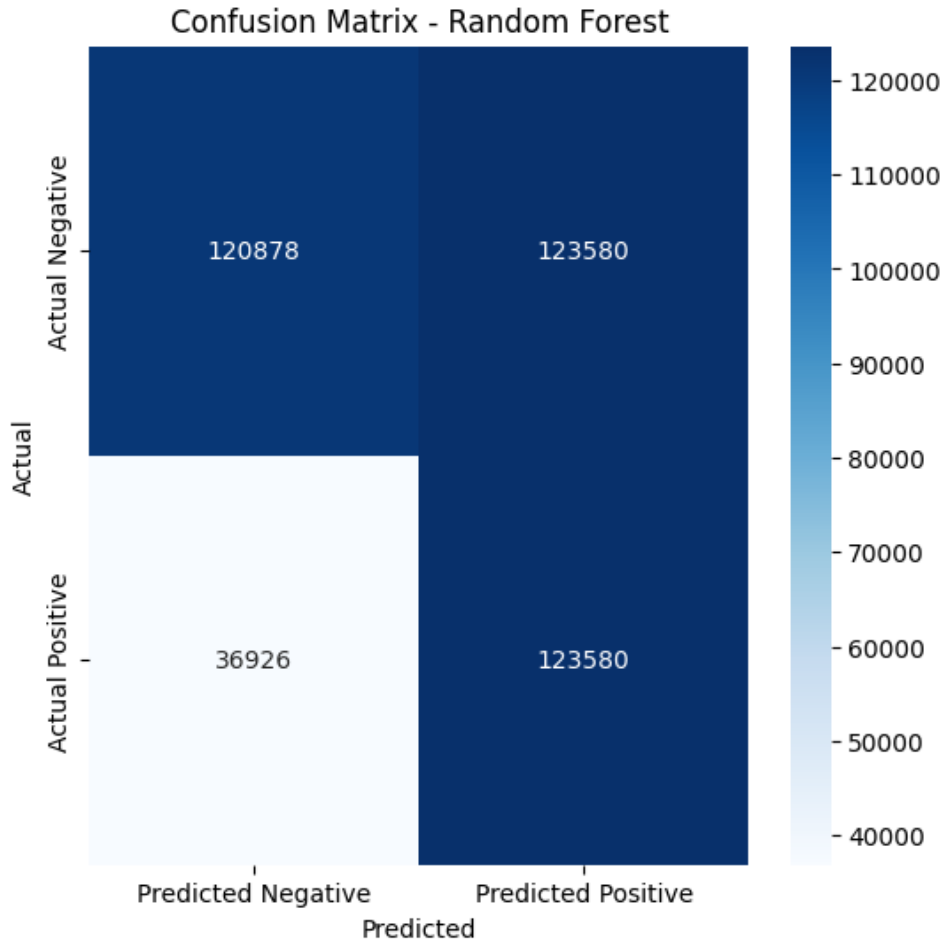

Figure 10: Random Forest

Interpretación de las Matrices de Confusión

Modelo XGBoost

- **True Negatives (Predicted Negative - Actual Negative)**: 9880
- **False Positives (Predicted Positive - Actual Negative)**: 13441
- **False Negatives (Predicted Negative - Actual Positive)**: 2681
- **True Positives (Predicted Positive - Actual Positive)**: 13441

La matriz de confusión para el modelo XGBoost muestra que el modelo tiene una cantidad significativa de falsos positivos (13441) y verdaderos negativos (9880). Esto indica que el modelo tiene dificultades para clasificar correctamente las muestras negativas.

Modelo Ridge

- **True Negatives (Predicted Negative - Actual Negative)**: 118340
- **False Positives (Predicted Positive - Actual Negative)**: 127959
- **False Negatives (Predicted Negative - Actual Positive)**: 32547

- **True Positives (Predicted Positive - Actual Positive)**: 127959

La matriz de confusión para el modelo Ridge muestra un alto número de falsos positivos (127959) y verdaderos negativos (118340). Aunque el modelo tiene un mejor desempeño en términos de verdaderos positivos, sigue teniendo una alta tasa de falsos positivos.

Modelo K-Nearest Neighbors (KNN)

- **True Negatives (Predicted Negative - Actual Negative)**: 12152
- **False Positives (Predicted Positive - Actual Negative)**: 7591
- **False Negatives (Predicted Negative - Actual Positive)**: 8531
- **True Positives (Predicted Positive - Actual Positive)**: 7591

La matriz de confusión para el modelo KNN muestra un balance entre los verdaderos negativos (12152) y los verdaderos positivos (7591), pero tiene un número considerable de falsos negativos (8531) y falsos positivos (7591).

Modelo Random Forest

- **True Negatives (Predicted Negative - Actual Negative)**: 120878
- **False Positives (Predicted Positive - Actual Negative)**: 123580
- **False Negatives (Predicted Negative - Actual Positive)**: 36926
- **True Positives (Predicted Positive - Actual Positive)**: 123580

La matriz de confusión para el modelo Random Forest muestra un alto número de verdaderos negativos (120878) y verdaderos positivos (123580), pero también tiene una cantidad considerable de falsos positivos (123580) y falsos negativos (36926).

*Conclusión

Al comparar las matrices de confusión de los diferentes modelos, se puede observar lo siguiente:

- El modelo **XGBoost** tiene una cantidad significativa de falsos positivos, lo que sugiere que clasifica incorrectamente muchas muestras negativas como positivas.
- El modelo **Ridge** también presenta un alto número de falsos positivos, aunque tiene un mejor desempeño en términos de verdaderos positivos.
- El modelo **K-Nearest Neighbors (KNN)** tiene un equilibrio entre verdaderos negativos y verdaderos positivos, pero con un alto número de falsos negativos.
- El modelo **Random Forest** tiene un alto número de verdaderos negativos y verdaderos positivos, pero también un alto número de falsos positivos y falsos negativos.

4 Conclusión

El análisis de sentimientos, implementado mediante algoritmos de clasificación, demuestra ser una herramienta eficaz para la optimización de campañas publicitarias. La capacidad de segmentar el mercado de manera más precisa y personalizar los mensajes publicitarios resulta en una mejora significativa en la efectividad y el retorno de inversión de las campañas. Futuras investigaciones pueden explorar la integración de estos métodos con otras técnicas de marketing digital para maximizar su impacto.

References

[1] Dalal, M. K., & Zaveri, M. A. (2011). Automatic Text Classification: A Technical Review. *International Journal of Computer Applications*, 28(2), 37-40.

[2] Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text Classification Using Machine Learning Techniques. *WSEAS Transactions on Computers*, 4(8), 966-974.

[3] Authors (2020). Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Tweets.

[4] Authors (2019). Sentiment Analysis Algorithms and Applications.

[5] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

[6] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.

[7] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12).

[8] Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)*, 34(1), 1-47.