



Análisis de sentimiento de los comentarios en TripAdvisor: un caso de los hoteles de Bucaramanga

Ángela Díaz Pinzón¹, Yuly Ramírez Sierra²

¹Facultad de Ingenierías Físico-mecánicas. Escuela de estudios Industriales y Empresariales. Ingeniería Industrial. Universidad Industrial de Santander, Colombia. angela2171116@correo.uis.edu.co

²Facultad de Ingenierías Físico-mecánicas. Escuela de estudios Industriales y Empresariales. Ingeniería Industrial. Universidad Industrial de Santander, Colombia. yaramsie@correo.uis.edu.co

Resumen

El análisis de sentimiento es una herramienta utilizada en el Procesamiento de Lenguaje Natural (PLN) para identificar el sentimiento o la emoción expresada en un texto. Es ampliamente utilizada para analizar el contenido generado por usuarios en diferentes campos. En este trabajo, se explora la aplicación del análisis de sentimiento en español al sector hotelero en la ciudad de Bucaramanga, extrayendo reseñas de seis de los hoteles mejor clasificados en TripAdvisor. Se trabaja con 350 reseñas, clasificándolas en tres polaridades: positivo, neutro y negativo, y separando cada reseña en dos componentes, los títulos y los contenidos. Se realiza el proceso de etiquetado manualmente y se toma como referencia para comparar con dos librerías multilingüe, VADER y Pysentimiento, así como con un modelo pre-entrenado basado en la arquitectura de transformadores, BERT, utilizando la validación cruzada para probar los hiperparámetros. Al comparar las dos librerías, se concluye que Pysentimiento funciona mejor para el problema de clasificación presentado. Con este resultado, se propone un esquema para el preprocesamiento de los datos, realizando combinaciones entre los pasos para medir la precisión, el recall y el F1-score de Pysentimiento. Se halla que, tanto para el escenario de los títulos como de los contenidos, funcionó mejor al aplicar uno de los pasos del preprocesamiento al conjunto de datos. Se entrena el modelo BERT y se obtienen resultados deficientes debido al tamaño del conjunto de datos y las clases desbalanceadas. Finalmente, se realiza un análisis para generar recomendaciones a los hoteles estudiados con base en los bigramas más frecuentes en cada polaridad detectada.

Palabras clave: Análisis de sentimiento, turismo Bucaramanga, hotelería Bucaramanga, minería de texto, comentarios en línea.

Abstract

Sentiment analysis is a tool used in Natural Language Processing (NLP) to identify the sentiment or emotion expressed in a text. It is widely used to analyze user-generated content in different fields. In this work, we explore the application of sentiment analysis in Spanish to the hotel sector in the city of Bucaramanga, extracting reviews from six of the top-rated hotels on TripAdvisor. We work with 350 reviews, classifying them into three polarities: positive, neutral, and negative, and separating each review into two components, titles and contents. The manual labeling process is carried out and used as a reference for comparison with two multilanguage libraries, VADER and Pysentimiento, as well as with a pretrained model based on the transformer architecture, BERT, using cross validation to test the hyperparameters. When comparing the two libraries, it is concluded that Pysentimiento performs better for the classification problem presented. With this result, a scheme for data preprocessing is proposed, performing

combinations between the steps to measure the precision, recall, and F1-score of Pysentimiento. It is found that, both for the titles and contents scenarios, it worked better to apply one of the preprocessing steps to the dataset. The BERT model is trained and poor results are obtained due to the size of the dataset and the imbalanced classes. Finally, an analysis is conducted to generate recommendations to the studied hotels based on the most frequent bigrams in each detected polarity.

Keywords: Sentiment analysis, Bucaramanga tourism, Bucaramanga hotel, text mining, online reviews.

1. Introducción

Actualmente, la cantidad de datos y contenido generado por usuarios está en casi todas las plataformas que se consumen. En los últimos años ha aumentado la tendencia de las personas a participar activamente en los diferentes espacios en línea, y también la tendencia a confiar en las reseñas o experiencias de los demás sobre todo al momento de adquirir un producto o un servicio, ya que las decisiones de compra se ven cada vez más influenciadas por las opiniones de terceros. Es por eso que surge la necesidad de las empresas de todos los sectores de centrar su atención hacia ese fenómeno, y con esto, buscar maneras para analizar la información brindada por los propios clientes de cierta manera. Con el avance del aprendizaje automático, crece no solo el interés sino también la facilidad de acceso a hacer uso de herramientas como el análisis de sentimiento para extraer información valiosa a partir de los datos, que no podrían ser procesados manualmente por su complejidad y cantidad.

En este contexto, el análisis de sentimiento emerge como una herramienta invaluable para las empresas, permitiéndoles entender las percepciones y emociones expresadas por los clientes en sus reseñas y comentarios en línea. A través del análisis de sentimiento, las organizaciones pueden identificar tendencias, evaluar la satisfacción del cliente, y tomar decisiones informadas para mejorar la calidad de sus productos y servicios.

En esta investigación, se explora el uso del análisis de sentimiento en el contexto del sector hotelero por medio de la plataforma TripAdvisor, centrado en la ciudad de Bucaramanga. Utilizando una combinación de técnicas de aprendizaje automático y herramientas de procesamiento de lenguaje natural, se busca comprender las percepciones de los clientes sobre los hoteles y generar recomendaciones basadas en la detección de patrones en los comentarios negativos, neutros y positivos.

2. Antecedentes y trabajos relacionados

Desde el año 2000 el análisis de sentimiento, también llamado minería de opinión, se ha convertido en una de las áreas más investigadas dentro del procesamiento de lenguaje natural, la minería de datos y la minería de texto, extendiéndose a las ciencias de gestión y las ciencias

sociales debido a su aplicación en distintas áreas de negocio [1]. Surge por la necesidad de analizar el sentimiento contenido de partes específicas de un texto, en lugar de clasificar un documento completo como bueno o malo [2].

El análisis de sentimiento hace referencia al estudio de las opiniones, sentimientos, actitudes y emociones hacia cierto tema y sus aspectos, y la detección de polaridad de dichas opiniones categorizándolas usualmente como positivas, negativas o neutras. Sin embargo, ha trascendido el problema de la detección de polaridad para despertar interés en la comunidad científica y en las empresas por su capacidad de predicción en áreas de finanzas y mercadeo [1]. El sector turismo no es una excepción a esta tendencia, especialmente porque el volumen de contenido y la facilidad de acceder a él han convertido a las reseñas en línea en un factor determinante en el proceso decisivo de los turistas, influenciando comportamientos como la intención de reservación y de visitar un hotel [3]. Los comentarios de los usuarios proveen una percepción más profunda sobre las preferencias de aspectos específicos de su estadía, por lo que ha surgido en la industria la necesidad de comprender sus fallas y corregir deficiencias para proveer un mejor servicio y satisfacer los requerimientos de los clientes [4]. Adicionalmente, es cada vez más común el uso de plataformas para realizar reservaciones como TripAdvisor, Booking.com, Airbnb y Yelp, la mayoría de estos sitios ofrecen también una sección de opiniones que da acceso a los usuarios a un gran volumen de información a la que pueden acudir para evaluar la calidad del hotel, y mitigar la incertidumbre que puede generar el visitar un lugar nuevo [5].

Como se indica en [6] los métodos de análisis pueden dividirse en tres categorías: RNN, CNN y Transformadores.

En [6] se usó un dataset etiquetado de reseñas en árabe y se usaron 8 métodos de aprendizaje profundo, dos de transformadores, 4 de RNN, 2 de CNN y una combinación entre RNN y CNN. Similar, en [7] se usaron Naive Bayes, máquinas de soporte vectorial, un modelo de CNN y dos de RNN con datos extraídos de TripAdvisor; en [8] Naive Bayes y árboles de decisión. En otros casos, como en [3], [9], [10], [11] no se enfocaron en probar modelos de clasificación, sino en las implicaciones para el sector de estudio al realizar el análisis de sentimiento.

Por último, la etiquetación manual se realiza en [12], [13], [14], [15], [16], [17], [18], [19], [20].

3. Metodología

La metodología del proyecto se describe en el esquema de Figura 1.

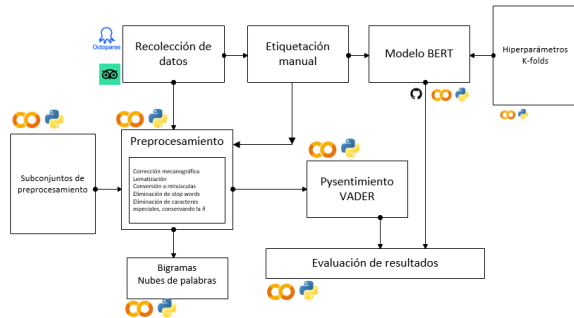


Figura 1. Esquema de la metodología del proyecto.
Fuente: elaboración propia.

En primer lugar, se realiza la recolección de datos desde la plataforma de TripAdvisor para seis hoteles en la ciudad de Bucaramanga: Hotel Dann Carlton, Hotel Hampton by Hilton, Hotel Barí, Holiday Inn Cacique, BGA Hotel y Hotel Chicamocha. El proceso de extracción de datos se realiza por medio del software Octoparse 8, especializado en la extracción masiva de datos desde páginas web. La herramienta utiliza un panel de operación visual para el que no se requiere codificación, lo que facilita el manejo. La extracción comienza ingresando el enlace de la página web de la que se descargarán los datos en el software, con lo que permite seleccionar mediante un clic los elementos que se desean extraer y crear un bucle para avanzar entre las páginas. A partir de esto, el software automáticamente extrae de la misma manera los datos de las demás reseñas. En la Figura 2 se presenta el flujo de trabajo.

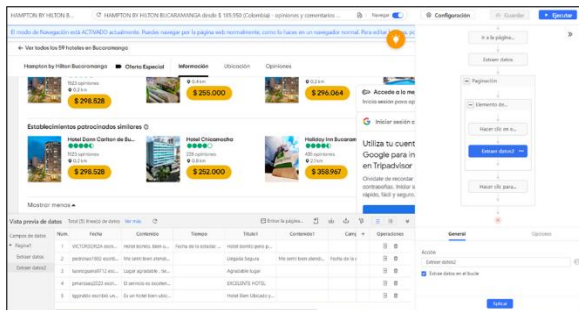


Figura 2. Flujo de trabajo en Octoparse.

Una vez finalizada la extracción, se exporta a un archivo en formato xlsx para cada hotel y se filtra por fecha de

publicación del comentario desde septiembre de 2020 hasta diciembre de 2023. Las reseñas incluyen el nombre del usuario acompañado de la fecha en la que fue escrita, la fecha de estadía, la calificación general en una escala de 1 a 5 por parte del usuario, el hotel correspondiente, el comentario dividido en dos componentes, el título del comentario y el contenido de este. Se obtienen en total 350 comentarios.

Se procede a revisar manualmente cada registro para asignar la etiqueta correspondiente del sentimiento: positivo, neutro o negativo. Si se resaltaban aspectos buenos sobre el hotel se asignaba la etiqueta de “Positivo”; si se presentaba a modo de queja o se notaba en general el disgusto en el comentario, se le asignaba la etiqueta “Negativo”; y se consideraba “Neutro” en dos ocasiones, primero, si el comentario no expresaba un sentimiento sino que daba alguna declaración que podría o no contener información específica sobre el hotel, o si mencionaba aspectos negativos y positivos en el mismo comentario de manera que se pudieran equilibrar y ninguno de los dos sentimientos predominara sobre el otro. Al finalizar el proceso de etiquetado se tiene para los títulos, 242 positivos, 56 neutros y 52 negativos, y para los contenidos, 246 positivos, 19 neutros y 85 negativos.

Una vez terminada la etiquetación manual, se realiza el preprocesamiento en Python, compuesto por 5 pasos: corrección mecanográfica, lematización, conversión a minúsculas, eliminación de stop words y eliminación de caracteres especiales conservando la letra ñ. Con esto se obtuvo un conjunto de nubes de palabras de bigramas por cada sentimiento.

Los hallazgos principales se describen a continuación. Para los comentarios negativos se mencionaron inconvenientes con el aire acondicionado (“aire acondicionado” y “habitación aire”), con la comida, el *check in*, la seguridad del hotel y el personal de recepción (ver Figura 3).

Para los comentarios clasificados como neutros, se encuentran algunos términos como “experiencia Dann”, “visita Bucaramanga”, “fin semana”, “primera vez”, y “booking.com”, que no proporcionan información sobre el sentimiento de los usuarios. Aunque, también se encuentran términos como “buena comida”, “buena atención” y “buen día”, que expresan algo positivo (ver Figura 4 y Figura 5). Esto sucede porque los comentarios no son completamente sesgados a un solo sentimiento, en este caso a pesar de mencionar algunos aspectos positivos es muy posible que en el mismo comentario se mencionaran cosas negativas que llevaron a determinar que se etiquetaran como neutros.



Figura 3. Nube de palabras de bigramas negativos para Contenido.



Figura 4. Nube de palabras de bigramas neutros para Título.



Figura 5. Nube de palabras de bigramas neutros para Contenido.

En los comentarios positivos resaltan aspectos como la ubicación del hotel, el servicio y la atención, el precio, la comida, y la limpieza y comodidad de las instalaciones: “buena ubicación”, “excelente servicio”, “excelente atención”, “calidad precio”, “comida deliciosa”, “habitación impecable”, “habitación cómoda” (ver Figura 6).



Figura 6. Nube de palabras de bigramas positivos.

Después, se comparan las dos librerías para análisis de sentimiento y se entrena el modelo BERT, tomando como referencia la etiquetación manual.

4. Experimentos y resultados

En este paso, se emplean las librerías de Pysentimiento y VADER para clasificar los comentarios en tres polaridades y comparar su desempeño con la etiquetación manual. Se utiliza la biblioteca de aprendizaje automático scikit-learn de Python para comparar los resultados y obtener matrices de confusión y métricas relevantes. Se presenta la matriz de confusión para los títulos identificados por VADER en la Figura 7, junto con las métricas en la Tabla 1. Los comentarios neutros muestran la menor precisión, pero el mayor recall. Esto sugiere que VADER clasificó muchos comentarios negativos y positivos incorrectamente como neutros (precisión), pero identificó muchos comentarios neutros correctamente en relación con los que debió haber clasificado (recall). Por otro lado, para los comentarios negativos, ocurre lo contrario: una alta precisión, pero el recall más bajo, indicando dificultad para reconocerlos. Finalmente, los comentarios positivos muestran la precisión y el F1-score más altos, indicando que fueron los mejor clasificados en general.

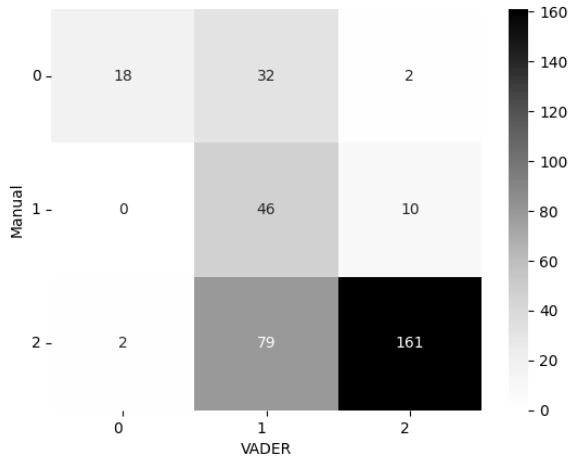


Figura 7. Matriz de confusión del etiquetado manual VS VADER para los títulos.

Los resultados para los contenidos se observan en la Figura 8 y la Tabla 2. Comenzando por los comentarios negativos, es evidente tanto en la matriz como en las métricas que no reconoció esta categoría, clasificando todos los comentarios como neutros y positivos. Para estas categorías, se tiene el mismo F1-score gracias a que sus medidas de precisión y recall son opuestas.

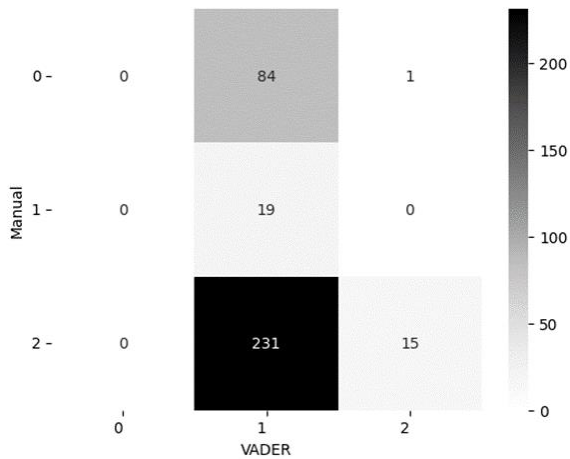


Figura 8. Matriz de confusión de etiquetado manual VS VADER para los contenidos.

Con Pysentimiento, se realizaron pruebas similares obteniendo las matrices de la Figura 9 (títulos) y la Figura 10(contenidos). En relación con los títulos, al analizar los comentarios negativos, se observan valores aceptables de precisión, recall y F1-score, como se detalla en la Tabla 1. En cuanto a los comentarios neutros, la precisión es baja, indicando que se clasificaron muchos comentarios erróneamente como positivos en comparación con los que se clasificaron correctamente como neutros. Por otro

lado, el recall es considerablemente mayor que el de los comentarios negativos, lo cual se debe a la clasificación de la mayoría de los neutros verdaderos como neutros, aunque también se clasificaron algunos como negativos o positivos. Sin embargo, el F1-score no es alto, posiblemente debido a la clasificación errónea de muchos positivos verdaderos como neutros. En cuanto a los comentarios positivos, se observa la precisión más alta, ya que Pysentimiento identificó la mayoría de manera correcta y solo una pequeña proporción como negativos o neutros. Aunque el recall fue el más bajo de las tres categorías, se clasificaron 88 comentarios como neutros cuando en realidad eran positivos según la etiquetación manual.

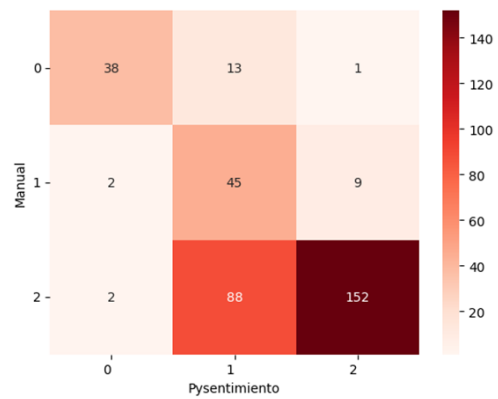


Figura 9. Matriz de confusión de etiquetado manual VS Pysentimiento para los títulos.

Para los contenidos, se lograron medidas altas de precisión, recall y F1-score para los comentarios negativos y positivos. Sin embargo, en cuanto a los comentarios neutros, el desempeño de las métricas es más bajo (ver Tabla 2). En este escenario, se identificaron correctamente la mayoría de los comentarios negativos en relación con el total clasificado como negativos y en comparación con los verdaderos negativos; un patrón similar se observó en los comentarios positivos.

En cuanto a los comentarios neutros, se clasificaron correctamente 8 comentarios como neutros, pero también se clasificaron 8 incorrectamente como negativos (precisión), aunque se logró clasificar la mayoría correctamente en relación con los neutros verdaderos (recall). Es importante señalar que el desempeño inferior en las tres métricas para los comentarios neutros posiblemente se deba a que solo 19 fueron etiquetados manualmente bajo esta categoría.

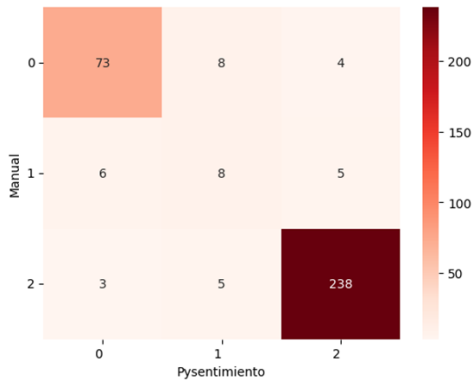


Figura 10. Matriz de confusión de etiquetado manual VS Pysentimiento para los contenidos.

Títulos						
Manual VS VADER	Precisión	Recall	F1-score	Support	Accuracy	MCC
Negativo	0.90	0.35	0.50	52	0.64	0.432
Neutro	0.29	0.82	0.43	56		
Positivo	0.93	0.67	0.78	242		
Manual VS Pysentimiento	Precisión	Recall	F1-score	Support	Accuracy	MCC
Negativo	0.90	0.73	0.81	52	0.67	0.501
Neutro	0.31	0.80	0.45	56		
Positivo	0.94	0.63	0.75	242		

Tabla 1. Métricas de los clasificadores para los títulos

Contenidos						
Manual VS VADER	Precisión	Recall	F1-score	Support	Accuracy	MCC
Negativo	0.0	0.0	0.0	85	0.1	0.07
Neutro	0.06	1.0	0.11	19		
Positivo	0.94	0.06	0.11	246		
Manual VS Pysentimiento	Precisión	Recall	F1-score	Support	Accuracy	MCC
Negativo	0.89	0.86	0.87	85	0.91	0.801
Neutro	0.38	0.42	0.40	19		
Positivo	0.96	0.97	0.97	246		

Tabla 2. Métricas de los clasificadores para los contenidos

Como segunda parte del experimento, al definir que Pysentimiento se comportó mejor, se aplicó el preprocesamiento al conjunto de datos y sus posibles combinaciones para probar si con este proceso mejora la clasificación de comentarios. Se obtuvo que clasifica mejor los títulos al eliminar de caracteres especiales, y mejor los contenidos al realizar la corrección mecanográfica.

Como último paso, se usó el modelo BERT para llevar a cabo el análisis de sentimiento. Se hace uso de un código de referencia proveniente del repositorio de Github denominado "Proyecto MAD", el cual está específicamente diseñado para el análisis de sentimiento multiclase en español, centrado en tweets relacionados con la Educación superior en Colombia [21]. El documento de entrada para el código es la etiquetación manual de los comentarios, donde se les asigna una

etiqueta numérica: 0 para negativo, 1 para neutro y 2 para positivo. Posteriormente, se procede a dividir el conjunto de datos, utilizando el 80% para el conjunto de entrenamiento y el 20% restante para el conjunto de prueba. Para la ejecución del modelo, se emplea la librería de scikit-learn y la función GridSearch para explorar las posibles combinaciones de hiper parámetros, tomando como referencia la Figura 11.

Hyperparameter	Values
Epochs	3, 4, 5
Batch Size	32
Learning Rate	2e-5, 3e-5, 5e-5, 6e-5, 7e-5, 8e-5, 1e-4
Weight Decay	0.1
Warmup Ratio	0.06, 0.08, 0.10

Figura 11. Hiper parámetros para los modelos considerados de Pysentimiento. Tomado de:[22].

Al ejecutar el modelo se obtuvo como resultado que, para el conjunto de entrenamiento, la mejor combinación incluye los siguientes parámetros:

Tasa de aprendizaje (lr): 3e-5

Épocas: 5

Warmup ratio (wr): 0.08

Para el conjunto de prueba, se observa la pérdida más baja en la última combinación, utilizando los siguientes hiperparámetros:

Tasa de aprendizaje (lr): 1e-4

Épocas: 5

Warmup ratio: 0.1

Finalmente, se evaluó el modelo con la combinación para el conjunto de entrenamiento.

Al observar la matriz de confusión (Figura 12) se evidencia que el modelo solo reconoció los comentarios positivos. Esto se debe a dos razones, principalmente, a la cantidad de datos que se usaron para entrenar el modelo, que probablemente no fueron suficientes para que este lograra identificar correctamente o reconocer cualquier otra categoría de comentario. En segundo lugar, las clases estaban desbalanceadas, lo que quiere decir que no había la misma cantidad de comentarios de cada categoría, al encontrarse con más comentarios positivos, fue la única categoría que el modelo logró identificar y clasificó todo bajo la misma.

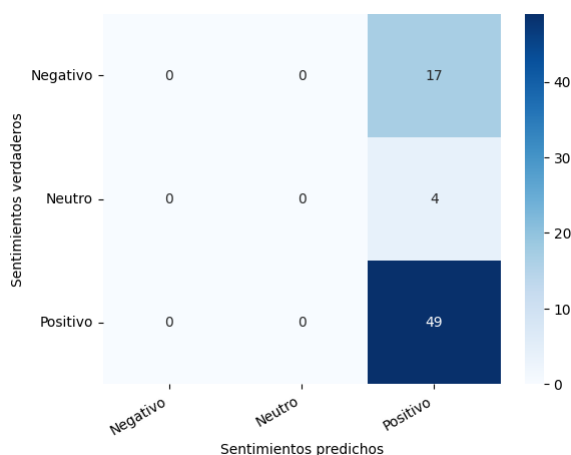


Figura 12. Matriz de confusión del etiquetado manual VS BERT para los contenidos.

Agradecimientos

Conflicto interés

Referencias

No es obligatorio usar los rótulos sugeridos, pero sí es preciso que en el artículo se incluyan los métodos o la metodología y los resultados, y que a partir de estos se generen unas conclusiones y recomendaciones. Se recomienda a los autores revisar [1], [2], [3], [4], [5] para tener una idea más clara de cómo escribir de manera apropiada un artículo científico original.

5. Conclusiones

Desde la revisión de literatura se ha destacado la importancia crucial del análisis de sentimiento como una herramienta invaluable para el sector hotelero. Este reconocimiento se fundamenta en la creciente disponibilidad de datos generados por usuarios en plataformas de viaje y alojamiento, así como en la capacidad del análisis de sentimiento para procesar eficientemente grandes conjuntos de datos. Estas plataformas almacenan una cantidad significativa de información que puede proporcionar conocimiento valioso sobre la satisfacción del cliente, las preferencias y las tendencias emergentes. El análisis de sentimiento se presenta como una solución para extraer información significativa de estos datos, permitiendo a las empresas del sector hotelero tomar decisiones informadas y estratégicas para mejorar la experiencia del cliente y la calidad del servicio basándose en los aspectos que son más importantes para los huéspedes.

En la extracción de datos desde TripAdvisor, aunque el proceso se limitó por cuestiones de disponibilidad de datos, se reconoce la importancia de esta base de datos

para comprender mejor las experiencias de los huéspedes en la ciudad. La limitación en el número de comentarios recopilados sugiere que podría haber una oportunidad para incentivar una mayor participación por parte de los hoteles y sus huéspedes en plataformas de reseñas en línea. Dado el papel significativo que juegan las opiniones de los usuarios en las decisiones de reserva y compra, esta iniciativa podría contribuir no solo a mejorar la visibilidad y reputación de los hoteles, sino también a promover el turismo en Bucaramanga, una ciudad que se posiciona como uno de los destinos turísticos emergentes en el país.

Durante la clasificación manual, dividiendo cada comentario en dos componentes, el enfoque permitió explorar si un título con cierto sentimiento necesariamente implicaba que todo el comentario compartiera ese mismo sentimiento, revelando que esta suposición no siempre se cumplía. Posteriormente, el esquema de preprocesamiento planteado en el proyecto permitió extraer información específica sobre los aspectos más frecuentes por cada sentimiento.

Luego, en la evaluación de VADER y Pysentimiento, determinando que Pysentimiento ofrecía un mejor rendimiento, se implementó el preprocesamiento dividiéndolo en pasos para evaluando todas las combinaciones posibles con el fin de experimentar su incidencia en la concordancia con la clasificación manual. Seguidamente, se utilizó el modelo BERT y se realizó una validación cruzada para ajustar los hiper parámetros, sin embargo, el rendimiento fue insatisfactorio debido a la limitada cantidad de comentarios y al desbalance de clases.

Finalmente, se realizaron análisis adicionales al centrarse por separado en los dos grupos de hoteles. Este análisis reveló que los comentarios negativos de los dos primeros hoteles se enfocaban principalmente en el aire acondicionado y el proceso de registro en el hotel, mientras que los neutros mencionaban la necesidad de aumentar la variedad, sin embargo, nada en concreto. En contraste, los comentarios positivos resaltaban el servicio del personal, llegando a la conclusión de que se valora este aspecto por encima de los demás. Para los otros cuatro hoteles, los comentarios negativos proporcionaron información más detallada, mencionando nuevamente el aire acondicionado y agregando el aspecto de las habitaciones y el servicio; mientras que los neutros indicaban problemas con la conexión wifi en uno de los hoteles. En los comentarios positivos, se valoraban aspectos como la ubicación y las instalaciones y la comida, aunque la atención al cliente no se consideraba tan relevante como en los dos primeros hoteles.

En resumen, los análisis revelan la importancia de considerar tanto la cantidad como la calidad de los comentarios al realizar el análisis de sentimiento, así como la necesidad de abordar el desbalance de clases y

la escasez de datos para obtener resultados más precisos y representativos.

6. Referencias

- [1] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan Claypool Publishers, 2012.
- [2] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the 2nd International Conference on Knowledge Capture, K-CAP 2003*, Association for Computing Machinery, Inc, Oct. 2003, pp. 70–77. doi: 10.1145/945645.945658.
- [3] G. Roy, "Travelers' online review on hotel performance – Analyzing facts with the Theory of Lodging and sentiment analysis," *Int J Hosp Manag*, vol. 111, May 2023, doi: 10.1016/j.ijhm.2023.103459.
- [4] J. Zhang, X. Lu, and D. Liu, "Deriving customer preferences for hotels based on aspect-level sentiment analysis of online reviews," *Electron Commer Res Appl*, vol. 49, Sep. 2021, doi: 10.1016/j.elerap.2021.101094.
- [5] R. Jayanto, R. Kusumaningrum, and A. Wibowo, "Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory," *International Journal of Advances in Intelligent Informatics*, vol. 8, no. 3, pp. 391–403, Nov. 2022, doi: 10.26555/ijain.v8i3.691.
- [6] R. A. Hameed, W. J. Abed, and A. T. Sadiq, "Evaluation of Hotel Performance with Sentiment Analysis by Deep Learning Techniques," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 9, pp. 70–87, 2023, doi: 10.3991/ijim.v17i09.38755.
- [7] K. Puh and M. Bagić Babac, "Predicting sentiment and rating of tourist reviews using machine learning," *Journal of Hospitality and Tourism Insights*, vol. 6, no. 3, pp. 1188–1204, Jun. 2023, doi: 10.1108/JHTI-02-2022-0078.
- [8] N. Z. Dina, "Tourist sentiment analysis on TripAdvisor using text mining: A case study using hotels in Ubud, Bali," 2020. [Online]. Available: <http://www.ajhtl.com>
- [9] İ. A. Özen and E. Özgül Katlav, "Aspect-based sentiment analysis on online customer reviews: a case study of technology-supported hotels," *Journal of Hospitality and Tourism Technology*, vol. 14, no. 2, pp. 102–120, Feb. 2023, doi: 10.1108/JHTT-12-2020-0319.
- [10] M. P. Mehta, G. Kumar, and M. Ramkumar, "Customer expectations in the hotel industry

- during the COVID-19 pandemic: a global perspective using sentiment analysis,” *Tourism Recreation Research*, vol. 48, no. 1, pp. 110–127, 2023, doi: 10.1080/02508281.2021.1894692.
- [11] D. Suryadi and J. A. Imran, “A Topic Modeling and Sentiment Analysis Approach for Benchmarking of Hotels Based on Online Reviews,” *Industrial Engineering and Management Systems*, vol. 21, no. 4, pp. 646–657, Dec. 2022, doi: 10.7232/iems.2022.21.4.646.
- [12] R. Sehgal, “Sentimental Analysis On Hotel Reviews Using Classic Approaches And A Smaller Data Set,” *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 8, no. 09, 2019, [Online]. Available: www.ijstr.org
- [13] D. N. Mishra and R. K. Panda, “How delightful is indian wellness tourism? A netnographic study,” *Advances in Hospitality and Tourism Research*, vol. 9, no. 1, pp. 132–156, 2021, doi: 10.30519/ahtr.784232.
- [14] N. Khamphakdee and P. Seresangtakul, “Sentiment analysis for Thai language in hotel domain using machine learning algorithms,” *Acta Informatica Pragensia*, vol. 10, no. 2, pp. 155–171, 2021, doi: 10.18267/j.aip.155.
- [15] T. Sontayasara *et al.*, “Twitter sentiment analysis of bangkok tourism during covid-19 pandemic using support vector machine algorithm,” *Journal of Disaster Research*, vol. 16, no. 1, pp. 24–30, 2021, doi: 10.20965/jdr.2021.p0024.
- [16] J. Luo, S. Huang, and R. Wang, “A fine-grained sentiment analysis of online guest reviews of economy hotels in China,” *Journal of Hospitality Marketing and Management*, vol. 30, no. 1, pp. 71–95, 2021, doi: 10.1080/19368623.2020.1772163.
- [17] S. Alosaimi, M. Alharthi, K. Alghamdi, T. Alsubait, and T. Alqurashi, “Sentiment analysis of arabic reviews for Saudi hotels using unsupervised machine learning,” *Journal of Computer Science*, vol. 16, no. 9, pp. 1258–1267, 2020, doi: 10.3844/jcssp.2020.1258.1267.
- [18] A. Benlahbib and E. H. Nfaoui, “Aggregating customer review attributes for online reputation generation,” *IEEE Access*, vol. 8, pp. 96550–96564, 2020, doi: 10.1109/ACCESS.2020.2996805.
- [19] T. S. Bang and V. Sornlertlamvanich, “Sentiment classification for hotel booking review based on sentence dependency structure and sub-opinion analysis,” *IEICE Trans Inf Syst*, vol. E101D, no. 4, pp. 909–916, Apr. 2018, doi: 10.1587/transinf.2016IIP0038.
- [20] C. H. Miranda, J. Guzmán, and D. Salcedo, “Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles,” *Procesamiento de Lenguaje Natural*, pp. 25–32, Mar. 2016.
- [21] E. Espinosa, M. Quiroga, and C. Rodriguez, “ProyectoMAD.” Oct. 2023. Accessed: Mar. 30, 2024. [Online]. Available: <https://github.com/moniiandrea/ProyectoMAD>
- [22] J. M. Pérez *et al.*, “pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks,” Jun. 2021, [Online]. Available: <http://arxiv.org/abs/2106.09462>