

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

Bucaramanga

Ángela María Díaz Pinzón

Trabajo de Grado para Optar al Título de Ingeniera Industrial

Director

Yuly Andrea Ramírez Sierra

Magister en Ingeniería Industrial

Universidad Industrial de Santander

Facultad de Ingenierías Físico-mecánicas

Escuela de Estudios Industriales y Empresariales

Ingeniería Industrial

Bucaramanga

2024

Dedicatoria

A la mejor persona del mundo, mi mamá, por su apoyo incondicional, al cual debo todo lo que soy.

Agradecimientos

A mi familia, por su infinito amor y comprensión.

A mis amigos, Laura, Andrea, Óscar, Valentina y Diego, que han sido parte fundamental en mi proceso de formación.

A mi directora Yuly, por su valioso acompañamiento, su paciencia y dedicación durante este proceso.

Tabla de Contenido

| | Pág. |
|---|-------------|
| Introducción | 13 |
| 1. Tabla de cumplimiento de objetivos | 15 |
| 2. Planteamiento del problema..... | 16 |
| 3. Objetivos | 20 |
| 3.1 Objetivo general | 20 |
| 3.2 Objetivos específicos | 20 |
| 4. Marco de referencia | 21 |
| 4.1 Marco de Antecedentes..... | 21 |
| 4.2 Marco Teórico..... | 25 |
| 4.2.1 Minería de datos..... | 25 |
| 4.2.2 Metodología KDD | 25 |
| 4.2.3 Minería de texto..... | 27 |
| 4.2.4 Análisis de sentimientos | 28 |
| 4.2.5 Aprendizaje profundo | 30 |
| 5. Revisión de la literatura | 35 |
| 5.1 Análisis Bibliométrico | 35 |
| 5.1.1 Publicaciones por año | 37 |
| 5.1.2 Publicaciones por país..... | 37 |
| 5.1.3 Publicaciones por área | 38 |
| 5.1.4 Publicaciones por autor | 40 |
| 5.1.5 Análisis de coocurrencia por palabras clave | 41 |

| | |
|---|-----|
| 5.2 Análisis preliminar de la literatura | 42 |
| 6. Proceso de descubrimiento del conocimiento en bases de datos | 48 |
| 6.1 Extracción y selección de datos | 48 |
| 6.2 Etiquetado manual del conjunto de datos | 55 |
| 6.3 Preprocesamiento de los datos | 59 |
| 6.4 Etiquetado de los datos con modelos pre-entrenados | 67 |
| 6.4.1 VADER | 70 |
| 6.4.2 Pysentimiento | 72 |
| 6.4.1 Evaluación de la detección de polaridad con Pysentimiento | 75 |
| 6.5 Análisis de sentimiento con BERT | 80 |
| 7. Utilidad del análisis de sentimientos en empresas hoteleras | 86 |
| 8. Conclusiones | 96 |
| 9. Limitaciones | 98 |
| 10. Recomendaciones | 99 |
| Referencias Bibliográficas | 100 |

Lista de Tablas

| | Pág. |
|---|-------------|
| Tabla 1. <i>Cumplimiento de objetivos</i> | 15 |
| Tabla 2. <i>Cantidad de comentarios por hotel</i> | 50 |
| Tabla 3. <i>Etiquetación manual de los comentarios</i> | 57 |
| Tabla 4. <i>Métricas de los clasificadores para los títulos.</i> | 75 |
| Tabla 5. <i>Métricas de los clasificadores para los contenidos</i> | 75 |
| Tabla 6. <i>Pasos del preprocesamiento de datos</i> | 76 |
| Tabla 7. <i>Resultados del coeficiente de correlación de Matthews de la clasificación de Pysentimiento con las posibles combinaciones de preprocesamiento de datos.</i> | 77 |
| Tabla 8. <i>Métricas de Pysentimiento para la clasificación de comentarios</i> | 80 |
| Tabla 9. <i>Hiper parámetros para los modelos considerados de Pysentimiento.</i> | 82 |
| Tabla 10. <i>Mejor combinación</i> | 83 |
| Tabla 11. <i>Métricas del modelo BERT para el análisis de sentimiento</i> | 86 |

Lista de Figuras

| | Pág. |
|--|-------------|
| Figura 1. <i>Red neuronal con tres capas (capa de entrada, capa oculta y capa de salida) y tres neuronas por capa.</i> | 31 |
| Figura 2. <i>Flujograma de ecuación de búsqueda</i> | 36 |
| Figura 3. <i>Número de publicaciones por año</i> | 37 |
| Figura 4. <i>Cantidad de publicaciones por país</i> | 38 |
| Figura 5. <i>Publicaciones por área</i> | 39 |
| Figura 6. <i>Relación de artículos entre principales áreas de conocimiento</i> | 40 |
| Figura 7. <i>Mapa de redes de coautoría presentado por VOSviewer</i> | 41 |
| Figura 8. <i>Mapa de redes de coocurrencia por palabras clave presentado por VOSviewer</i> | 42 |
| Figura 9. <i>Herramientas para análisis de sentimiento en los artículos de la revisión de literatura.</i> | 46 |
| Figura 10. <i>Modelos de aprendizaje automático en los artículos de la revisión de literatura.</i> | 47 |
| Figura 11. <i>Cantidad de comentarios por hotel</i> | 50 |
| Figura 12. <i>Selección de la estructura de la reseña en Octoparse</i> | 51 |
| Figura 13. <i>Selección de los datos dentro de la reseña</i> | 52 |
| Figura 14. <i>Interfaz de trabajo de Octoparse</i> | 52 |
| Figura 15. <i>Panel de sugerencias de Octoparse</i> | 53 |
| Figura 16. <i>Flujo de trabajo en Octoparse</i> | 54 |
| Figura 17. <i>Proceso de limpieza</i> | 60 |
| Figura 18. <i>Nube de palabras de bigramas negativos para Título.</i> | 63 |

| | |
|---|----|
| Figura 19. <i>Nube de palabras de bigramas negativos para Contenido.</i> | 63 |
| Figura 20. <i>Nube de palabras de bigramas neutros para Título</i> | 64 |
| Figura 21. <i>Nube de palabras de bigramas neutros para Contenido</i> | 65 |
| Figura 22. <i>Nube de palabras de bigramas positivos para Título</i> | 66 |
| Figura 23. <i>Nube de palabras de bigramas positivos para Contenido.</i> | 66 |
| Figura 24. <i>Fragmento del arrchivo de Excel con la clasificación de los comentarios por Pysentimiento.</i> | 68 |
| Figura 25. <i>Fragmento del archivo de Excel con la clasificación de los comentarios por VADER.</i> | 68 |
| Figura 26. <i>Matriz de confusión del etiquetado manual VS VADER para los títulos.</i> | 70 |
| Figura 27. <i>Matriz de confusión de etiquetado manual VS VADER para los contenidos.</i> | 71 |
| Figura 28. <i>Matriz de confusión de etiquetado manual VS Pysentimiento para los títulos.</i> | 73 |
| Figura 29. <i>Matriz de confusión de etiquetado manual VS Pysentimiento para los contenidos.</i> ... | 74 |
| Figura 30. <i>Matriz de confusión de etiquetado manual VS Pysentimiento para los títulos después de eliminar caracteres especiales en el conjunto de datos.</i> | 79 |
| Figura 31. <i>Matriz de confusión de etiquetado manual VS Pysentimiento para los contenidos después de realizar la corrección mecanográfica al conjunto de datos.</i> | 80 |
| Figura 32. <i>Resultados de los parámetros</i> | 82 |
| Figura 33. <i>Métricas de los conjuntos de entrenamiento y prueba.</i> | 84 |
| Figura 34. <i>Matriz de confusión del etiquetado manual VS BERT para los contenidos</i> | 85 |
| Figura 35. <i>Gráfica de barras de la cantidad de comentarios por categoría del primer grupo de hoteles</i> | 87 |

| | |
|---|----|
| Figura 36. <i>Gráfica de barras del porcentaje de comentarios por categoría del primer grupo de hoteles</i> | 87 |
| Figura 37. <i>Nube de palabras y diagrama de frecuencias de comentarios negativos para el primer grupo</i> | 88 |
| Figura 38. <i>Nube de palabras y diagrama de frecuencias de comentarios neutros para el primer grupo.</i> | 89 |
| Figura 39. <i>Nube de palabras y diagrama de frecuencias de comentarios positivos para el primer grupo.</i> | 90 |
| Figura 40. <i>Gráfica de barras de la cantidad de comentarios por categoría del segundo grupo de hoteles</i> | 91 |
| Figura 41. <i>Gráfica de barras del porcentaje de comentarios por categoría del segundo grupo de hoteles</i> | 92 |
| Figura 42. <i>Nube de palabra y diagrama de frecuencias de comentarios negativos para el segundo grupo de hoteles</i> | 93 |
| Figura 43. <i>Nube de palabras y diagrama de frecuencias de comentarios neutros para el segundo grupo de hoteles</i> | 94 |
| Figura 44. <i>Nube de palabras y diagrama de frecuencias de comentarios positivos para el segundo grupo de hoteles.</i> | 95 |
| Figura 45. <i>Esquema de la metodología del proyecto</i> | 95 |

Lista de Apéndices

Apéndice A. Librerías y modelos encontrados en la revisión de literatura.

Apéndice B. Cantidad de comentarios por hotel obtenidos desde TripAdvisor.

Apéndice C. Componentes de los comentarios extraídos de TripAdvisor.

Apéndice D. Código para preprocesamiento de comentarios y creación de nubes de palabras.

Apéndice E. Clasificación de comentarios por VADER.

Apéndice F. Clasificación de comentarios por Pysentimiento.

Nota: Los apéndices se encuentran adjuntos en una carpeta

Resumen

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

Autor: Ángela María Díaz Pinzón**

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

Descripción: 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 multilinguaje, VADER y Pysentimiento, así como con un modelo pre entrenado basado en la arquitectura de transformadores, BERT, utilizando la validación cruzada para probar los hiper parámetros. 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 un análisis para generar recomendaciones a los hoteles estudiados con base en los bigramas más frecuentes en cada polaridad detectada.

* Trabajo de Grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Estudios Industriales y Empresariales. Ingeniería Industrial. Director: Yuly Andrea Ramírez Sierra. Magíster en Ingeniería Industrial

Abstract

Title: Sentiment analysis of TripAdvisor reviews: a case of Bucaramanga hotels*

Author(s): Ángela María Díaz Pinzón**

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

Description: 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.

* Degree Work

**Faculty of Physical-Mechanical Engineering. School of Industrial and Business Studies. Industrial Engineering. Director: Yuly Andrea Ramírez Sierra. Master's in Industrial Engineering.

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. 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.

En esta investigación se detalla el marco de antecedentes, las teorías y conceptos que soportan el tema objeto de estudio, la revisión de literatura con los hallazgos más importantes, el análisis bibliométrico, el proceso de descubrimiento de conocimiento de bases de datos (por sus siglas en inglés, KDD) aplicado, las utilidades del análisis de sentimiento para el sector hotelero, las conclusiones, limitaciones, recomendaciones, referencias bibliográficas y los apéndices.

1. Tabla de cumplimiento de objetivos

Tabla 1. *Cumplimiento de objetivos*

| OBJETIVO | CUMPLIMIENTO |
|--|--|
| 1. Realizar una revisión de literatura sobre el análisis de sentimientos de los comentarios en línea asociados al sector turismo y/o al sector hotelero. | Error! Reference source not found. , capítulo 5 |
| 2. Consolidar la base de datos con las opiniones que los clientes de los hoteles de Bucaramanga han realizado en Tripadvisor. | Error! Reference source not found. Apéndice C Error! Reference source not found. |
| 3. Clasificar la información recolectada a partir de los sentimientos expresados por parte de los clientes de los hoteles objeto de estudio. | Error! Reference source not found. Error! Reference source not found. Análisis de sentimiento con BERT Utilidad del análisis de sentimiento |
| 4. Elaborar un artículo de carácter publicable a partir de los resultados de la investigación con el fin de identificar tendencias de la percepción de los clientes respecto al servicio recibido. | |

2. Planteamiento del problema

Colombia es el tercer país de Latinoamérica con mejor recuperación del sector turístico tras la crisis por el COVID-19 y el confinamiento, excediendo durante el 2022 en 1,7% las cifras del 2019 (Ministerio de Comercio, Industria y Turismo, 2023). De acuerdo con la Organización Mundial del Turismo de las Naciones Unidas (OMT), el turismo contribuye a la reducción de las desigualdades territoriales a través de la promoción del desarrollo económico local en destinos estratégicos (Ministerio de Comercio, Industria y Turismo, 2022), por lo que la planificación y gestión adecuadas influyen en la mejora de la calidad de vida, la protección del patrimonio cultural y los recursos naturales, el desarrollo de infraestructuras y la reanimación del comercio en las ciudades (Ministerio de Comercio, 2022). El Plan Sectorial de Turismo identifica que los principales retos del sector en el país son la generación de estrategias de publicidad para la comercialización de los servicios de los hoteles, la capacidad para dotar de sentido a un destino y la sincronización con la tecnología, y resalta la importancia de fortalecer las herramientas de calidad y fácil acceso con el fin de elaborar estudios estadísticos que encaminen al fortalecimiento empresarial y la productividad del sector (Ministerio de Comercio, Industria y Turismo, 2022).

Para el área metropolitana de Bucaramanga, la ocupación hotelera alcanzó una tasa de 55,14% durante el 2022, superando al 2019 en un 5,98%. Durante el mismo año la tasa de ocupación de la ciudad fue de 60,4%, representando el sector turismo el 24,26% de la totalidad y el subsector de alojamiento 156.495 empleados, es decir, el 27,7% del sector turismo (Instituto Municipal de Cultura y Turismo de Bucaramanga, 2022). Asimismo, durante los primeros cinco meses del 2023 se registró la llegada de 335.445 pasajeros al Aeropuerto Internacional Palonegro en vuelos nacionales y 40.298 en vuelos internacionales. De igual forma, se identifica que tanto

los turistas extranjeros como nacionales, escogen como tipo de alojamiento los hoteles sobre los hostales o el alojamiento con algún familiar o amigo (Instituto Municipal de Cultura y Turismo de Bucaramanga, 2022) .

Adicionalmente, con el avance de la tecnología y la facilidad que implica para los turistas utilizar plataformas de reservación en línea por la variedad de opciones, la rapidez y la posibilidad de consultar opiniones de los aspectos importantes del hotel, las experiencias compartidas por los usuarios se han convertido en una fuente de información que impacta la reputación de los prestadores del servicio e influye en la toma de decisiones del turista (Alaei et al., 2019). Es muy común acudir a información generada por otros usuarios antes de realizar reservas de estadía, ya que los usuarios se han convertido en los protagonistas de la búsqueda de información, incrementando su participación directa en las plataformas de viaje mediante la creación y divulgación de contenido (Zaida De León Martín et al., 2020).

Una de las plataformas más populares actualmente entre los turistas es Tripadvisor, que permite consultar comentarios sobre alojamientos, restaurantes o aerolíneas, comparar precios y hacer reservaciones; actualmente cuenta con 859 millones de opiniones disponibles en 28 idiomas (Equipo de administración Tripadvisor, 2019), tiene 490 millones de visitantes mensualmente y recibe 700.000 contribuciones de reseñas por hora (Boshnjakoska, 2023). Según Similar Web, en Colombia es la página web más visitada en la categoría de Viajes y Turismo, seguida por despegar.com y pricetravel.com (Similar web, 2024).

Es evidente que estas plataformas almacenan gran cantidad de datos, con lo que surge la oportunidad de realizar análisis utilizando distintos métodos con el fin de descubrir conocimiento y apoyar la toma de decisiones de los interesados. En este sentido, los investigadores consideran la minería de texto como un recurso ideal para extraer la información de los canales digitales (Das

et al., 2022), desarrollar modelos de análisis del valor turístico, construir sistemas de recomendación turística, crear perfiles turísticos y formular políticas para supervisar los mercados turísticos. La popularidad del método se debe también a los avances del procesamiento del lenguaje natural (PLN), el aprendizaje automático y el aprendizaje profundo (Li et al., 2019). La minería de texto trabaja con datos semiestructurados o no estructurados y se encarga de extraer información indicada explícitamente en el texto a analizar. Los métodos más comunes para analizar los datos no estructurados extraídos de diferentes de fuentes son: resumen, extracción de información, categorización, visualización, agrupación, seguimiento de temas y análisis de sentimiento (Kaushik & Naithani, 2016).

Particularmente, el Análisis de Sentimientos (AS) se define como el procesamiento de mensajes directos, reseñas o comentarios en línea sobre un tema específico con el fin de extraer opiniones, emociones o actitudes de los usuarios y clasificarlos según su polaridad (positivo, negativo, neutro). Además, reúne investigaciones de informática, minería de datos, recuperación de texto y lingüística computacional, por lo que ofrece oportunidades de aplicación muy amplias tanto para trabajo cuantitativo como cualitativo (Mejova, 2009). Debido a su complejidad y la cantidad de datos existentes, el tipo de información que puede obtenerse abarca desde los sentimientos de los clientes sobre determinado producto o servicio hasta las opiniones públicas sobre situaciones sociales o políticas (Jain et al., 2021). Por lo tanto, sus aplicaciones son de gran potencial, permitiendo la predicción de resultados electorales, de comportamientos del consumidor, la evaluación de nuevos productos u ofertas, la elección de estrategias de mercadeo, entre otras (Mejova, 2009).

De esta manera, el presente proyecto tiene como objetivo realizar un análisis de sentimientos de las reseñas de los hoteles mejor calificados en la plataforma de Tripadvisor en la

ciudad de Bucaramanga, con el fin de comprender el sentimiento de los turistas hacia el servicio recibido. En este contexto, el análisis de sentimientos se presenta como una herramienta valiosa para los prestadores de servicios turísticos de la ciudad de Bucaramanga ya que permite obtener información precisa sobre la experiencia de los usuarios al considerar sus opiniones sobre el servicio, identificar si sus necesidades y expectativas fueron satisfechas, visualizar en términos generales el desempeño del hotel e identificar si existen patrones en los sentimientos hacia aspectos específicos del hotel, con el fin de apoyar la toma de decisiones en los procesos de gestión de los establecimientos con base la información brindada por los turistas.

3. Objetivos

3.1 Objetivo general

Analizar los sentimientos en los comentarios de los principales hoteles de Bucaramanga disponibles en TripAdvisor con el fin de identificar tendencias en la percepción de los clientes respecto al servicio recibido.

3.2 Objetivos específicos

Realizar una revisión de literatura sobre el análisis de sentimientos de los comentarios en línea asociados al sector turismo y/o al sector hotelero.

Consolidar la base de datos con las opiniones que los clientes de los hoteles de Bucaramanga han realizado en Tripadvisor.

Clasificar la información recolectada a partir de los sentimientos expresados por parte de los clientes de los hoteles objeto de estudio.

Elaborar un artículo de carácter publicable a partir de los resultados de la investigación con el fin de identificar tendencias en los clientes respecto al servicio recibido.

4. Marco de referencia

4.1 Marco de Antecedentes

Para la construcción del marco de antecedentes se tomaron como referencia tres proyectos de grado con enfoque en análisis de sentimientos, desarrollados en Colombia, Chile y Portugal. A continuación, se presenta una descripción de los proyectos, la metodología utilizada, los resultados obtenidos y la relación con el presente proyecto.

En primer lugar, (Méndez Pineda & Talero Sarmiento, 2021) plantean un modelo que permita predecir el movimiento de las principales acciones de la bolsa colombiana (Ecopetrol, Índice Colcap y Bancolombia) diaria y semanalmente, aplicando el aprendizaje automático a los datos históricos de los índices más representativos. En el análisis de la literatura y el marco teórico se exponen generalidades de métodos predictivos sin hacer énfasis en el análisis de sentimiento, posteriormente se describen algunos métodos y diccionarios útiles para el entrenamiento del conjunto de datos durante el análisis de sentimiento. Para la construcción del modelo, se toma como fuente de datos no estructurados el portal de noticias colombiano La República y realiza un Web crawling de las noticias para aplicar el análisis de sentimientos, utilizando Procesamiento de lenguaje natural (PLN) y los clasificadores: Textblob, VADER, SenticNet y LoughMcDonald. Se clasifican 60 noticias sobre los indicadores económicos del país por medio de los cuatro diccionarios en dos modalidades con un rango de -2 a 2 que permite asignar un valor numérico según el sentimiento de la noticia (muy positivo, positivo, neutro, negativo o muy negativo). Este proceso se realizó con cuatro diccionarios, 2 modalidades y 3 indicadores, con lo que se obtienen 24 matrices de confusión y del mismo modo las diferentes comparaciones que permitieron llegar a conclusiones sobre el desempeño de cada diccionario para cada indicador.

Después de realizar la clasificación de las noticias se compara con la serie de retorno de las acciones del día en el que fue publicada la noticia. Finalmente, se concluye que según el porcentaje de precisión no puede establecer que el sentimiento de las noticias nacionales tenga un impacto positivo sobre las acciones nacionales. De este proyecto pueden tomarse como referencia los métodos y algoritmos de clasificación explorados en el análisis preliminar de la literatura; la importancia y la metodología del adecuado preprocesamiento (selección y limpieza) de los datos, al igual que la investigación sobre las herramientas y diccionarios usados para la clasificación debido a que su desempeño puede variar significativamente según el contexto en el que sean aplicados y ciertas palabras pueden tener una connotación incorrecta por el mismo motivo; y las recomendaciones sobre la relación entre el contexto y el desempeño.

Por otro lado, (Gárce Chaparro, 2019) plantea la construcción de un software que permita la extracción automatizada de información en español desde Twitter, la procese, realice el análisis de sentimiento y entregue un resumen de lo encontrado. En primer lugar, se realiza Web Scraping para la extracción de tweets y se emplea un bot que los descarga en Python 3, junto con tres bibliotecas, Geckodriver, phantomjs y BeautifulSoup. Después de la descarga se realiza una limpieza de texto que permita utilizar la información como la base de datos. Se procede al entrenamiento con la biblioteca de Scikit learn de Python, que incluye varios algoritmos de clasificación, regresión y análisis, este proceso se divide en cuatro pasos que son: etiquetado, entrenamiento, pruebas y despliegue de resultados. Para el análisis de sentimiento se usan todos los tweets seleccionados y el muestreo se lleva a cabo por medio de la biblioteca xlswriter que genera automáticamente archivos de Excel con distintas funciones, en este caso se analizan 30 unigramas, 30 bigramas y 30 trigramas con mayor frecuencia para cada empresa. Este proceso se realiza para los tuits de los usuarios de las empresas Claro Chile y Enel Chile en los últimos cuatro

meses del 2018. El entrenamiento se lleva a cabo con cuatro clasificadores, SVM, regresión logística, árbol de decisión y Naive Bayes Multinomial, y se establecen las etiquetas según lo encontrado en los comentarios de los usuarios y la naturaleza del servicio de las empresas. Para cada clasificador se realizan 7 iteraciones, y se expone un análisis de desempeño de cada uno basándose en las métricas mencionada en el marco teórico. Se llega a la conclusión que el mejor clasificador para las dos empresas objeto de estudio fue Regresión Logística, y finalmente, se presentan los principales problemas identificados en las empresas según la percepción de los usuarios, basándose en el análisis de sentimientos realizado, y las recomendaciones respecto a las limitaciones en los procesos de aprendizaje automático, análisis de sentimientos y preprocesamiento del texto.

Este proyecto aporta la conceptualización y ejemplificación clara y concisa de términos fundamentales del análisis de sentimiento tales como los métodos del aprendizaje automático, métodos de clasificación y bibliotecas de uso abierto. Brinda una explicación detallada de la metodología para el proceso de entrenamiento y realiza una serie de recomendaciones para cada punto del proceso, al igual que alternativas de librerías para dar solución a las limitaciones presentadas en el análisis de sentimiento. Adicionalmente, ofrece un análisis de desempeño de cada clasificador que facilita la comprensión de las métricas de rendimiento, las limitaciones y recomendaciones de cada uno, al igual que la comparación entre ellos. Al mismo tiempo, un análisis de los problemas más frecuentes de cada empresa, lo que da un enfoque administrativo que puede ser útil al momento de las recomendaciones para el sector hotelero.

Finalmente, (T. Yu, 2020) realiza un análisis de sentimiento de las opiniones en línea de los usuarios sobre los restaurantes de Giethoorn. Toma como objeto de estudio el restaurante Zwaantje para compararlo con los mejores calificados en las reseñas y proponer mejoras para dicho

restaurante. Por medio de Lexalytics, una herramienta de inteligencia artificial de minería de texto que permite descomponer las reseñas en elementos gramaticales al tiempo que identifica el contexto de estos elementos, se analizan 4.832 opiniones hasta finales de agosto de 2018 de 28 restaurantes desde Google, Tripadvisor y Facebook en cuatro categorías: calidad de alimentos y bebidas, servicio, ambiente y valor. Se escogen entre 20 y 100 reseñas para cada restaurante, descartando de este modo los que tengan menos de 20 entradas, y tomando las más recientes para aquellos que tengan más de 100. Se comienza con la recolección de las reseñas y se procede a clasificarlas manualmente en Lexalytics en las cuatro categorías establecidas, de esta manera la herramienta extrae el sentimiento de cada reseña y realiza la comparación de desempeño de los restaurantes. Se genera un análisis de las cuatro categorías para los restaurantes con mejor desempeño en cada una y el restaurante Zwaantje, en este se encuentra que para todas las categorías prevalecen las opiniones neutrales, sin embargo, el porcentaje de opiniones negativas para Zwaantje es mucho mayor al de los demás. Al concluir la investigación se sugieren aspectos en los que la administración del restaurante debería enfocarse para mejorar el rendimiento, basándose en las palabras con mayor frecuencia en los comentarios.

A diferencia de los primeros proyectos, este tiene un enfoque turístico, que reafirma las ventajas de la aplicación del análisis de sentimientos a esta área, comprobando cómo puede llevar a la mejora en la toma de decisiones gerenciales a través del reconocimiento de las percepciones de los consumidores, debido a la cantidad de comentarios en línea y a la importancia que se les da cada vez más a las experiencias de otros usuarios al momento de tomar una decisión. Su principal aporte se encuentra en el reconocimiento de las limitaciones al momento de realizar el proyecto, ya que de esta manera se hace énfasis en la importancia de tener en cuenta aspectos como la adaptación de las bibliotecas o librerías al contexto de la investigación, los elementos gramáticos

y semánticos como los intensificadores y las negaciones que pueden afectar o distorsionar en mayor o menor medida la identificación de la polaridad del sentimiento expresado, y las diferencias en el proceso al tratarse de un idioma diferente al inglés. Adicionalmente, brinda un marco de trabajo de referencia y da la visión de un plan de acción para proponer mejoras al servicio prestado, lo que puede tenerse en cuenta en las recomendaciones del proyecto.

4.2 Marco Teórico

4.2.1 Minería de datos

La minería de datos responde a la necesidad de analizar la basta cantidad de datos disponibles en casi todos los aspectos de la cotidianidad. Cada día se registran millones de datos relacionados con gran variedad de ámbitos como transacciones de ventas, retroalimentación de los clientes, vigilancia ambiental, registros médicos, producción de contenido audiovisual, etc. ya sea vía web o en los distintos dispositivos de almacenamiento. La minería de datos consiste en el uso de herramientas que automatizan el proceso de convertir imaginables cantidades de datos en información organizada que pueda ser útil para la toma de decisiones (Han et al., 2012). Se conoce también como “descubrimiento del conocimiento”, o en inglés Knowledge Discoverey in Databases (por sus siglas en inglés, KDD), aunque se considera en otros casos como el paso fundamental de este proceso, que se describirá a continuación.

4.2.2 Metodología KDD

Esta metodología hace referencia al proceso de descubrimiento de conocimiento en bases de datos y busca extraer información previamente desconocida y potencialmente útil de datos sin procesar (Fayyad, Piatetsky-Shapiro, Smyth & Uthurusamy, 1996, como se citó en (Hernandez-

Orallo, 2005)). Los patrones o tendencias extraídos del texto pueden utilizarse para predecir comportamientos en el contexto al que se esté aplicando. Consiste en la secuencia interactiva de los siguientes pasos:

a. Integración de datos: es la unión de datos provenientes de varias fuentes que permite reducir redundancias e inconsistencias. Al momento de emparejar los datos se debe prestar mucha atención a la estructura para asegurarse de que lo que se tiene en una base de datos es equivalente a lo que está en otras.

b. Selección de datos: consiste en la aplicación de técnicas de reducción para obtener y seleccionar una representación más pequeña de los datos que sea relevante para el análisis que se va a realizar sin sacrificar la integridad de los datos originales. Las estrategias más comunes son la reducción de dimensionalidad, la reducción de numerosidad y la compresión de datos.

c. Limpieza de datos: comienza con la identificación de discrepancias en los datos e incluye remover datos con ruido e inconsistentes, completar valores faltantes y resolver inconsistencias. Como es muy probable que los datos inicialmente carezcan de precisión, la ausencia de la limpieza puede crear desconfianza en los resultados de cualquier extracción de datos posterior.

d. Transformación de datos: en este punto se consolidan los datos en el formato apropiado sin sacrificar la integridad de ellos, en ocasiones puede realizarse antes de la selección de datos. Influye en la eficiencia de la minería de datos y en la facilidad para comprender los patrones extraídos.

e. Minería de datos: se aplican métodos para extraer patrones de los datos.

f. Evaluación de patrones: es la identificación de patrones importantes, asociaciones y relaciones entre los datos que se alinean con el análisis de interés. Es muy útil para estrategias de marketing, análisis de decisiones y gestión empresarial. Dependiendo del contexto puede acudirse

a estrategias más avanzadas para el descubrimiento de patrones, como las asociaciones multinivel, multidimensionales, cuantitativas y los patrones raros, negativos, de alta dimensión, comprimidos y aproximados.

g. Presentación del conocimiento: exponer el conocimiento extraído mediante técnicas de visualización y representación.

Los cuatro primeros pasos corresponden a diferentes formas de preprocesamiento de los datos en el que se preparan para la minería, que pueden mejorar notablemente la calidad de los patrones encontrados y el tiempo requerido para extraerlos. Las razones para realizar el preprocesamiento se relacionan con la precisión, integridad, coherencia, credibilidad e interpretabilidad de los datos, que generalmente faltan en las grandes bases de datos reales.

A pesar de que la metodología se encuentra definida para toda clase de datos que puedan ser objeto de interés para la extracción de información, existen algunos tipos de datos que implican mayor complejidad por su naturaleza, por ejemplo, datos multimedia, de texto, datos espaciales y espaciotemporales (Han et al., 2012).

4.2.3 Minería de texto

Es un campo interdisciplinario basado en la recuperación de información, la minería de datos, el aprendizaje automático y los métodos estadísticos. Usualmente requiere estructurar un texto de entrada, seguido por la derivación de patrones de los datos estructurados y la evaluación e interpretación de la salida. Las tareas más comunes de la minería de texto son la categorización y agrupación de texto, la extracción de conceptos, el análisis de sentimiento y el resumen de documentos (Han et al., 2012).

4.2.4 Análisis de sentimientos

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 (Liu, 2012). 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 (Nasukawa & Yi, 2003). 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 (Liu, 2012). Puede por ejemplo mejorar la gestión de las relaciones con los clientes o los sistemas de recomendación para descubrir qué características gustan a los clientes y qué puede excluirse de las recomendaciones al haber recibido en su mayoría comentarios negativos; igualmente, se aplica en estrategias de marketing para recopilar y predecir actitudes del público hacia productos y marcas específicas; por otro lado, es útil para la detección de spam y el filtrado de comentarios ofensivos, irrelevantes o repetitivos con el fin de mantener un ambiente en línea más seguro. Existen tres enfoques principales para abordar el análisis de sentimiento: técnicas basadas en el conocimiento, métodos estadísticos y enfoques híbridos.

Las técnicas basadas en conocimiento clasifican el texto en categorías por medio de la presencia de palabras afectivas evidentes como “feliz”, “triste”, “amor”, “enojado”, “miedo” o “dolor”. La principal debilidad del enfoque es el reconocimiento equivocado del sentimiento al

estar involucradas reglas lingüísticas como la negación, por ejemplo, identificando erróneamente como feliz la frase “hoy no me siento feliz en absoluto”.

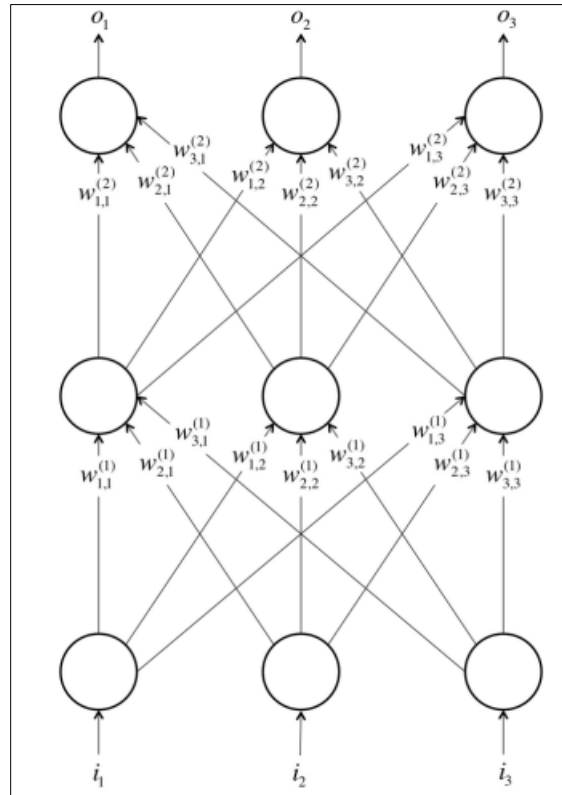
Por su parte los métodos estadísticos, como las máquinas de vectores de soporte y el aprendizaje profundo, funcionan alimentando un algoritmo de aprendizaje automático con un corpus de entrenamiento para que el sistema aprenda a reconocer la polaridad de las palabras afectivas y la frecuencia de la coexistencia de las palabras. Sin embargo, presentan dificultad para comprender el significado y la relación semántica entre las palabras, por lo que los elementos que aparecen juntos con frecuencia pueden tener poco valor predictivo al ser considerados individualmente, ya que el hecho de que ciertas palabras aparezcan juntas con frecuencia no quiere decir que tengan una relación semántica específica entre sí. Esto lleva a que los métodos estadísticos solo se comporten con precisión al proporcionar un conjunto de texto de entrada lo suficientemente grande.

Finalmente, los enfoques híbridos explotan los dos primeros enfoques con el objetivo de comprender mejor las reglas conceptuales del sentimiento y los indicios de estas reglas para entender la percepción, el procesamiento y la expresión de las emociones humanas. Últimamente, estos enfoques se han encargado de establecer el análisis de sentimiento como un campo interdisciplinario para abordar el estudio de las emociones desde múltiples perspectivas, evolucionando de las técnicas basadas en la sintaxis o la estructura gramatical del texto hacia estrategias semánticas o conscientes del significado de las palabras y su contexto (Cambria et al., 2017)

4.2.5 Aprendizaje profundo

El aprendizaje profundo es un subconjunto del aprendizaje automático, que es a su vez un campo de la inteligencia artificial. El aprendizaje automático se basa en la idea de aprender a partir de ejemplos, se le da a una computadora un modelo que le permita evaluar ejemplos y un conjunto de instrucciones que modifique el modelo cuando se equivoque. Para la construcción de este modelo se necesita plantear una estructura, tener unos parámetros y asignar los valores más adecuados a estos parámetros, lo que se convierte en un problema de mayor complejidad al tener datos que son poco estructurados y más difíciles de comprender y procesar, como es el caso del texto. Para enfrentarse a la complejidad los modelos más recientes de aprendizaje automático buscan recrear la estructura del cerebro humano en la construcción de dichos modelos, a esto se le llama aprendizaje profundo. La unidad fundamental del cerebro es la neurona, cuya funcionalidad y estructura se traducen a un modelo artificial que se pueda representar en una computadora, re expresando su funcionalidad en forma de vector. Como la transmisión de información no es posible con una sola neurona, tanto para el cerebro humano como para el aprendizaje profundo, se construyen redes neuronales artificiales (**Error! Reference source not found.**). Los métodos de aprendizaje profundo superan los de aprendizaje automático y llegan incluso a estar a la par con la precisión humana en problemas de visión por computadora y procesamiento del lenguaje natural (Buduma & Locascio, 2017).

Figura 1. Red neuronal con tres capas (capa de entrada, capa oculta y capa de salida) y tres neuronas por capa.



Nota. Tomado de Buduma y Locascio (2017, p.10)

A partir de las técnicas que se han desarrollado para la construcción de las redes neuronales surgen los métodos de aprendizaje profundo, los principales se describen brevemente a continuación.

a. Redes neuronales convolucionales (CNN): su origen se remonta a 1998 con el artículo “Gradient-based learning applied to document recognition”, de Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Las redes convolucionales utilizan tres ideas básicas, los campos receptivos locales, los pesos compartidos y la agrupación (Nielsen, 2015). El funcionamiento de

una capa convolucional consiste en procesar un volumen tridimensional de información para producir un nuevo volumen tridimensional de información.

Primero, una capa convolucional toma un volumen de entrada con ancho, alto, profundidad y relleno cero. Este volumen se procesa por una cantidad k de filtros de convolución con ciertos hiper parámetros que representa los pesos y las conexiones de la red, los filtros se aplican a regiones locales de los datos de entrada para extraer características (Buduma & Locascio, 2017). La red aprende automáticamente los filtros para identificar patrones utilizando una operación secuencial en la que la salida de una capa es la entrada de la siguiente, de esta manera se agrupan características para obtener representaciones jerárquicas de los datos (Hameed et al., 2023). Suelen ser más efectivas para identificar patrones locales en datos espaciales y pueden presentar dificultad para manejar relaciones en datos secuenciales, por eso su aplicación en reconocimiento de imágenes, detección de objetos y segmentación semántica.

b. Redes neuronales recurrentes (RNN): la idea general de las redes recurrentes existe cierta noción de cambio dinámico en el tiempo (Nielsen, 2015). Al igual que las redes convolucionales utiliza datos secuenciales, pero su diferencia radica en que las redes recurrentes comparten parámetros en las capas de la red y su memoria almacena información de entradas anteriores que influencia tanto la entrada como la salida siguiente. Todas las neuronas tienen conexiones entrantes que provienen de todas las neuronas de la capa anterior y conexiones de salida que conducen a todas las neuronas a la siguiente capa, a su vez, tienen conexiones recurrentes que transmiten información a neuronas de la misma capa, por lo que una capa recurrente con r neuronas tendrá r^2 conexiones recurrentes (Buduma & Locascio, 2017).

c. Transformadores: se propone por primera vez en (Vaswani et al., 2017). Es el primer modelo de transducción de secuencias basado en atención. La arquitectura del modelo depende

completamente de un mecanismo de autoatención de múltiples cabezas para generar dependencias globales entre la entrada y la salida, evitando la recurrencia. El método se basa en el entrenamiento de dos redes neuronales, la primera genera un vector para identificar las características que se consideran interesantes, y la segunda genera la descripción de características para identificar una relación. Las palabras son transformadas en vectores y de esta manera logra identificar a qué parte de la frase se está prestando atención. Su principal ventaja es que resuelve el problema de la memoria a corto plazo de las redes convolucionales y recurrentes, contextualizando cada palabra con las demás de la oración sin importar qué tan lejos estén debido a que aprende dependencias de largo alcance en la red. Para esto se considera la longitud del camino que se debe recorrer entre dependencias hacia delante y hacia atrás, cuanto más cortos sean estos caminos entre cualquier combinación de posiciones en las secuencias de entrada y salida, más fácil será aprender dependencias de largo alcance (Vaswani et al., 2017).

d. Bidirectional Encoder Representation from Transformers (BERT): es un modelo basado en transformadores creado para pre entrenar representaciones de texto no etiquetado teniendo en cuenta el contexto en dos direcciones, tanto de derecha a izquierda como de izquierda a derecha. El modelo se introdujo en 2019 para la respuesta a preguntas y la traducción de texto, mostrando un desempeño superior al de otros en los puntajes GLUE, MultiNLI y SQuAD, sin embargo, puede ser usado para una gran variedad de tareas gracias a su arquitectura que se divide en preentrenamiento y ajuste. Primero, se pre entrena el modelo en diferentes tareas con datos no etiquetados, para después ajustar todos los parámetros del preentrenamiento usando datos etiquetados de las tareas secundarias. Es importante resaltar que cada tarea tiene distintos modelos de ajuste a pesar de iniciar con los mismos parámetros de preentrenamiento.

Las representaciones de entrada están compuestas de un token inicial de clasificación llamado CLS, la secuencia de entrada y el token de separación entre oraciones llamado SEP. El preentrenamiento se realiza con 330 millones de palabras de BooksCorpus y Wikipedia y consiste en dos tareas, Masked LM y Next Sentence Prediction. La primera se basa en ocultar o enmascarar algunas palabras del corpus de texto con el fin de que el modelo prediga estas palabras correctamente; y la segunda consiste en predecir la siguiente oración de un texto escogiendo dos oraciones de entrada A y B, donde la mitad de las veces es B la oración que sigue a A y la otra mitad es alguna oración aleatoria dentro del corpus. De esta manera, las dos tareas ayudan a BERT a capturar información contextual y a aprender representaciones de palabras y oraciones contextualizadas que son muy útiles para las tareas de PLN (Devlin et al., 2019).

5. Revisión de la literatura

5.1 Análisis Bibliométrico

El análisis bibliométrico se lleva a cabo con el fin de analizar la literatura existente por medio de una búsqueda estructurada en la base de datos Scopus. En primer lugar, se seleccionan las palabras clave: “Sentiment analysis”, “Hospitality and tourism”, “Text mining”, “Online comments”. A partir de estas se determina una serie de sinónimos que permitieron la generación de la siguiente ecuación de búsqueda:

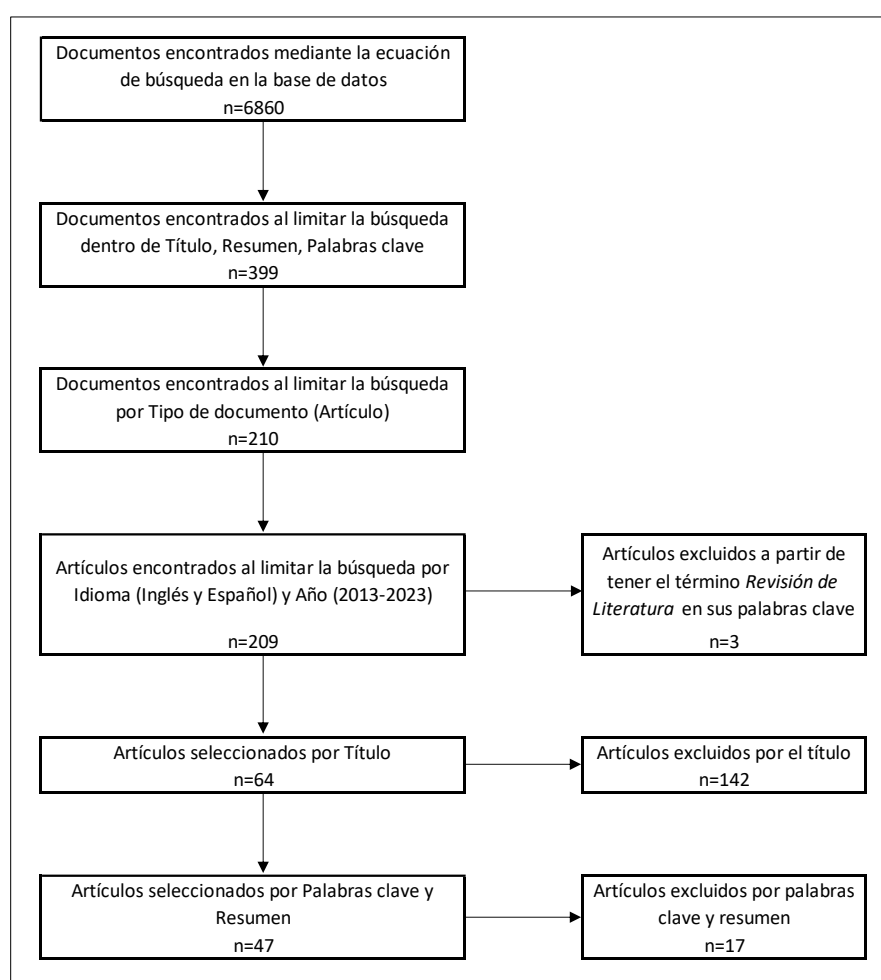
```
(TITLE-ABS-KEY ("sentiment analysis" OR "data mining") AND TITLE-ABS-KEY ("online comment" OR "social media" OR tripadvisor OR booking OR "online review") AND TITLE-ABS-KEY (hotel* OR lodging))
```

Inicialmente, la ecuación arroja 6.860 documentos, de los cuales se descartan 6.461 después de limitar la búsqueda por título, palabras clave o resumen, es decir, que los términos introducidos en la ecuación estuvieran contenidos en alguno de estos tres componentes de los documentos. Así, con 399 artículos se procede a limitar la búsqueda por año de publicación desde 2013 hasta 2023, por idioma a español e inglés, y por tipo de documento a artículo, con lo que son preseleccionados 206. Finalmente, se realiza la revisión por título, palabras clave y resumen, para determinar qué artículos abordaban aplicaciones del análisis de sentimiento o minería de opinión al sector de hotelería y/o turismo, descartando los artículos que no tenían mayor relación con el objeto de esta investigación. Se identifican tres patrones principales en el contenido que permitieron excluir gran cantidad de artículos. En la mayoría de los casos el análisis de sentimiento se enfoca al sector turístico u hotelero; en otros, por el contrario, se centra en aspectos específicos del sector sin aplicar análisis de sentimientos; por último, el análisis de sentimiento no es el núcleo

central de la investigación sino que se utiliza como uno de los pasos para completar un objetivo diferente, por ejemplo, para la construcción de un software de recomendación de hoteles el artículo no dedicaría una explicación detallada al análisis de sentimientos.

De manera que el total de artículos incluidos dentro de la revisión de literatura es de 47, excluyendo el 77,18% tal como se indica en la Figura 2.

Figura 2. Flujograma de ecuación de búsqueda

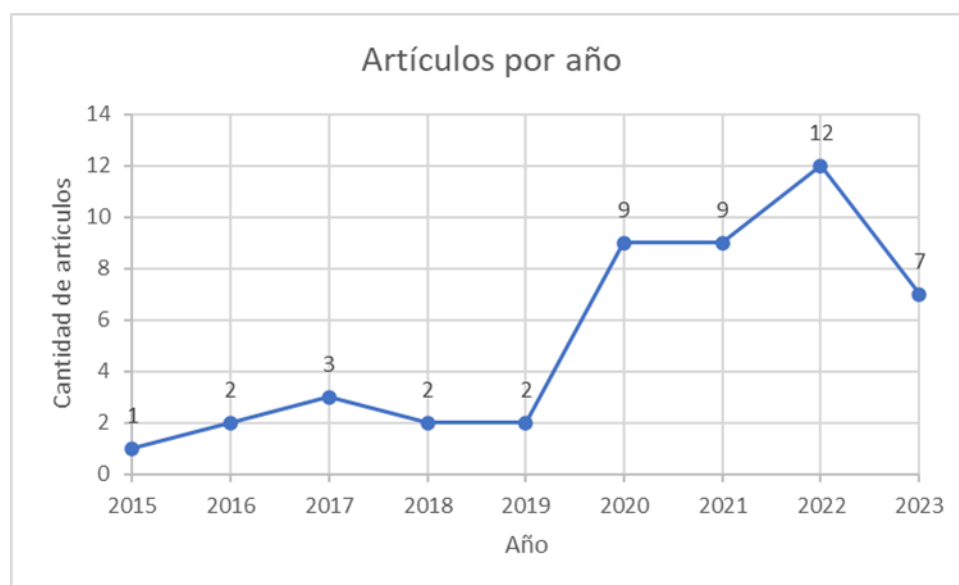


De esta manera, se procedió a realizar el análisis bibliométrico de los documentos por medio de Scopus y VosViwer.

5.1.1 Publicaciones por año

En la Figura 3 se observa que la mayoría de los artículos pertenecen a los últimos cuatro años. Al filtrar la búsqueda de 2013 a 2023, no se obtienen resultados para el 2023 ni 2014, y hasta el 2019 se encuentran tan solo 10 publicaciones, que no superan 3 por año. Se presenta un aumento para el 2020, con una cantidad de 9 que se mantiene constante para el año posterior hasta alcanzar el máximo de 12 artículos por año en el 2022, finalizando con 7 para la primera mitad del 2023.

Figura 3. Número de publicaciones por año



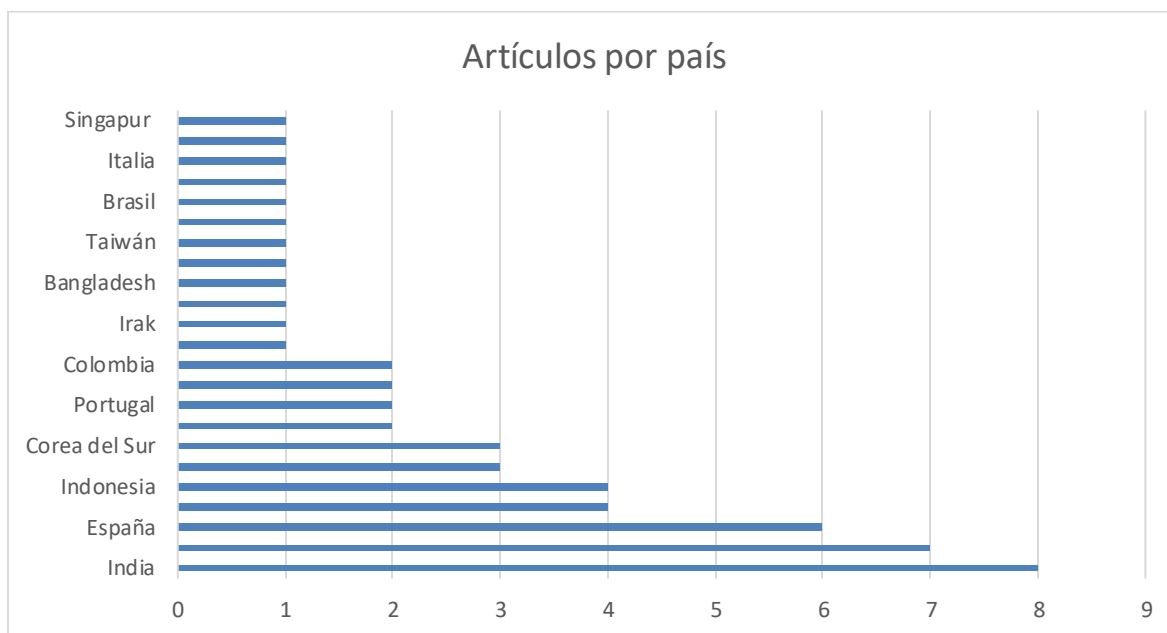
Nota. Adaptado de Scopus.

5.1.2 Publicaciones por país

Como se evidencia en la Figura 4, el 58,18% de los artículos seleccionados pertenecen a países asiáticos, encabezados por India con 8 publicaciones y China con 7, seguidos por España con 6 y los demás países europeos que corresponden aproximadamente a un cuarto (25,45%) de las publicaciones. Estados Unidos y Colombia son los dos países de América con más de una

publicación, representando cerca del 11% con respecto al total. Dentro de los países que tienen un artículo se encuentran Turquía, Croacia, Brasil, Canadá, Singapur, Taiwán e Italia.

Figura 4. Cantidad de publicaciones por país



Nota. Adaptado de Scopus.

5.1.3 Publicaciones por área

Se encuentran trece áreas de conocimiento para los artículos seleccionados. Más del 80% de los artículos seleccionados se relacionan con cuatro de las trece áreas de conocimiento, como se puede observar en la Figura 5. Se encuentran 23 documentos dentro de Ciencias de la computación, 22 dentro de Negocios, gestión y contabilidad, 14 dentro de Ciencias Sociales y 8 dentro de Ingeniería. Cada documento puede clasificar dentro de más de un área de conocimiento, se identifica que para las cuatro mencionadas, 6 artículos tienen en común Ciencias de la computación e Ingeniería, mientras que Negocios, gestión y contabilidad comparte 9 artículos con Ciencias sociales, sin embargo, no se encuentra ninguno que pertenezca a las cuatro áreas al

tiempo. Esto se debe a que gran cantidad de los artículos se basa en el análisis de sentimientos con el fin de conocer las opiniones y motivaciones de los usuarios y proponer mejoras al servicio brindado por los lugares turísticos para los que se realizó el estudio, lo que hace referencia al área de Negocios y Ciencias sociales; casi en igual proporción se exploran técnicas o herramientas de procesamiento de lenguaje natural y se proponen modelos para mejorar la clasificación de la polaridad de las opiniones, haciendo referencia a las Ciencias de computación e Ingeniería. Las relaciones entre las áreas de conocimiento principales se pueden observar en la Figura 6.

Figura 5. *Publicaciones por área*

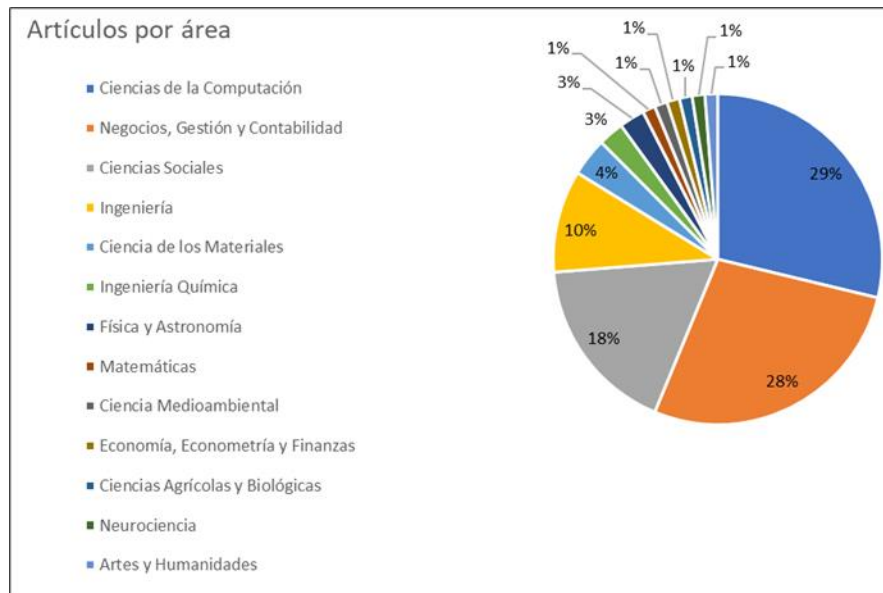
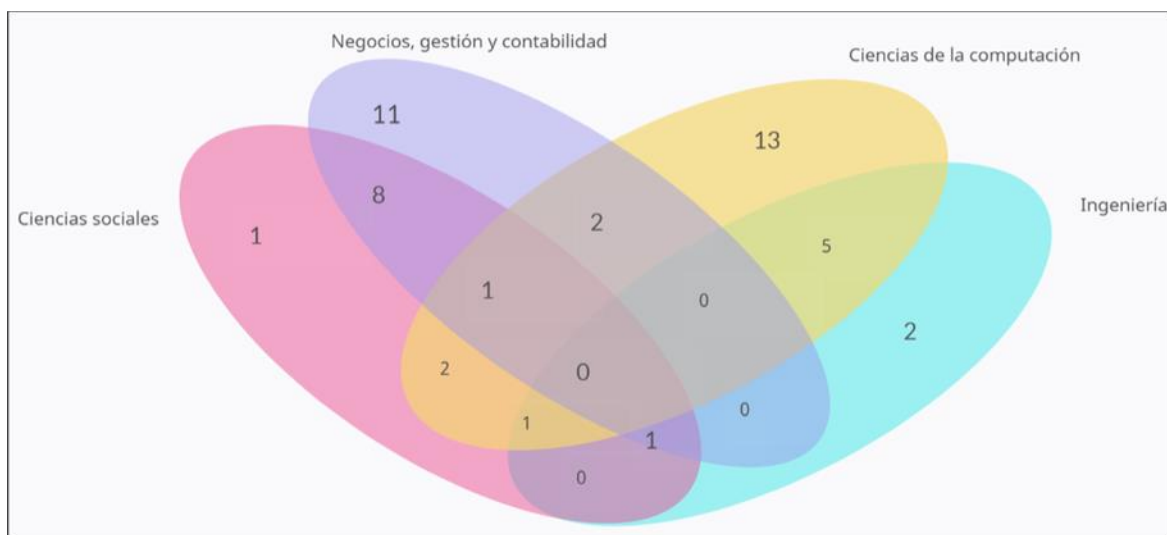


Figura 6. Relación de artículos entre principales áreas de conocimiento

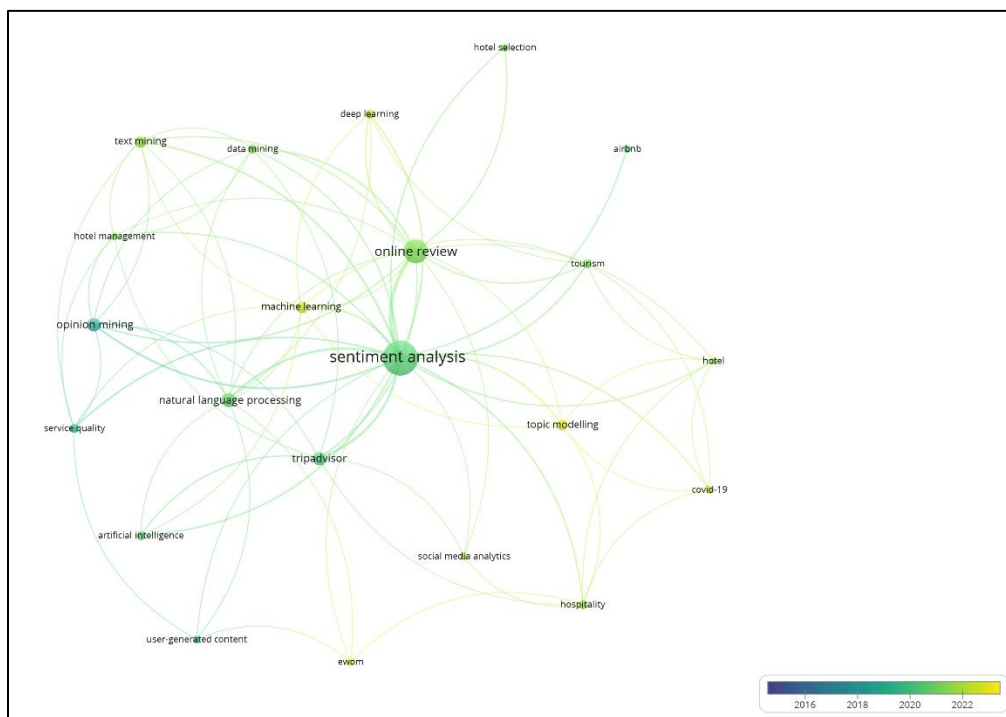
Nota. Adaptado de Scopus. **Error! Reference source not found.**

5.1.4 Publicaciones por autor

Para los 47 artículos se encuentra un total de 152 autores, la mayoría de ellos tienen un solo artículo publicado, exceptuando a Herrera F., Luzón M.V., Valdivia A. y Law R., que cuentan con dos artículos cada uno. Respecto a las citas, Geetha M., Singha P. y Sinha S. son los autores más citados, con 206, seguidos por Herrera F., Luzón M.V. y Valdivia A., con 166, y Jeong M., Lee J., y Lee M. con 146. Para la generación de la **Error! Reference source not found.**, con el objetivo de observar las relaciones de coautoría, se seleccionan un mínimo de 14 citas por autor teniendo en cuenta que los valores anteriores a este involucraban autores que no compartían autoría en los documentos, con lo que se obtuvo 61 autores relacionados mediante 15 redes, la mayoría de ellas compuestas por cuatro autores o menos. Law R. comparte autoría con 7, conformados por dos redes de 4 y 3 autores. Por otro lado, se encuentran dos redes de 7 autores conformadas por Cambria E., Chaturvedi I., Herrera F., Hrabova E., Luzón M.V., Troiano L. y

demás términos de la figura, seguido por ‘*online review*’ con 19 apariciones y relacionado con 14 términos, entre estos ‘*natural language processing*’ que aparece en 7 documentos y ‘*tripadvisor*’ y ‘*opinion mining*’ que aparecen en 6. Después se encuentran ‘*data mining*’, ‘*hospitality*’ y ‘*tourism*’ en 3 de los documentos, cada uno relacionado con 7 términos. Se destaca la aparición de términos como el boca a boca electrónico, ‘*ewom*’ en 2 artículos y ‘*topic modelling*’ en 4, a partir del 2022.

Figura 8. Mapa de redes de coocurrencia por palabras clave presentado por VOSviewer



5.2 Análisis preliminar de la literatura

Para el análisis preliminar de la literatura se revisan los artículos seleccionados, identificando en ellos la tendencia de los usuarios a generar contenido en las redes y confiar en las reseñas disponibles al momento de tomar una decisión sobre su estadía, al igual que la relevancia

del análisis de sentimiento para el sector turístico y hotelero, y los distintos modelos usados para clasificar el sentimiento.

Durante los últimos años el contenido generado por usuarios ha tomado gran importancia debido a la creciente popularidad de las redes sociales y los sitios de reseñas que reciben miles de opiniones cada día. 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 (Roy, 2023). 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 (Zhang et al., 2021). 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 (Jayanto et al., 2022). Por lo tanto, dichas plataformas se utilizan cada vez más como fuente de información para los turistas, las empresas e investigadores que desean ahondar en el tema de las reseñas en línea, sus motivaciones e implicaciones.

La razón de la importancia y el creciente interés en la aplicación del análisis de sentimiento se debe a la oportunidad de comprender de manera más amplia los puntos de vista de los usuarios, tarea que es posible por los métodos de análisis automatizados, debido a que no sería viable procesarlas manualmente por la cantidad y variedad. Por medio de los resultados del análisis, los hoteles pueden descubrir el nivel de importancia que dan los huéspedes a distintos factores,

identificar aspectos que deben priorizarse y necesidades que no han logrado satisfacer (Suryadi & Imran, 2022) o comparar el desempeño con el de sus competidores (Oliveira et al., 2022a). Basados en esto pueden tomar decisiones que mejoren notablemente la calidad del servicio, ya que pueden enfocarse en los aspectos más valiosos para los turistas, explorar la implementación de estrategias, e incluso predecir tasas de ocupación y tendencias de demanda.

Una investigación basada en el análisis de sentimiento puede conducir a mejorar la toma de decisiones gerenciales que se vean reflejadas en las métricas de rendimiento, así como a la predicción del desempeño del hotel según distintos indicadores, por ejemplo, en términos de ocupación (Chang et al., 2021) o ganancias (Nicolau et al., 2023), y en contextos más específicos analizar la percepción de los clientes sobre aspectos como el uso de la tecnología (Özen & Özgül Katlav, 2023), lo que permite comprender en qué enfocar sus esfuerzos, o si ciertas categorías son irrelevantes para los usuarios.

Como se indica en (Hameed et al., 2023) los métodos de análisis pueden dividirse en tres categorías: RNN, CNN y Transformadores. Un gran número de publicaciones se basa en comparar el rendimiento entre varios modelos (Imtiaz et al., 2021) (Puh & Bagić Babac, 2023) (Hameed et al., 2023) (Nicolau et al., 2023) (Arroni et al., 2023) (Ounacer et al., n.d.) (Chang et al., 2021), sin embargo, no se puede llegar a una conclusión respecto al más recomendable debido a que, dependiendo del contexto de la investigación, el conjunto de datos para el entrenamiento y las librerías empleadas, los mismos métodos pueden obtener rendimientos que varían significativamente. Si se usa un conjunto de imágenes para entrenar un modelo, es probable que no demuestre buenos resultados al aplicarlo para un análisis de sentimiento de comentarios en línea. Por lo tanto, no se trata de argumentar sobre si los clasificadores son buenos o malos, sino de buscar el conjunto de datos adecuado, etiquetarlo y entrenar el modelo correctamente. Se

encuentra que únicamente los transformadores coinciden en tener mejor desempeño en cada uno de los estudios, dado que estos resuelven el problema de otros métodos de memoria a corto plazo y permiten relacionar cada palabra de la oración sin tener en cuenta la distancia entre estas.

Las herramientas utilizadas en los artículos revisados para el análisis de sentimiento, etiquetación de datos o detección de polaridad se presentan en la **Error! Reference source not found.** Se muestra que la etiquetación manual fue la más común, aplicándose en nueve artículos (Mishra & Panda, 2021), (Khamphakdee & Seresangtakul, 2021) (Sontayasara et al., 2021) (J. Luo et al., 2021) (Alosaimi et al., 2020) (Benlahbib & Nfaoui, 2020) (Bang & Sornlertlamvanich, 2018) (Miranda et al., 2016). Todos comparten la característica de realizar análisis de sentimiento en comentarios escritos en idiomas diferentes al inglés, incluyendo español, árabe, tailandés y vietnamita. En segundo lugar, se encuentra VADER con 6 artículos (Roy, 2023), (Ounacer et al., n.d.), (Suryadi & Imran, 2022), (Imtiaz et al., 2021), (Win et al., 2022), (J. M. Luo et al., 2021), (Valdivia et al., 2019), seguido por el uso de conjuntos de datos que ya estuvieran etiquetados (Lee et al., 2017) (Sehgal, 2019) (Dina, 2020) (Puh & Bagić Babac, 2023) (Hameed et al., 2023). En esta última categoría se incluyen los procesos en los que se tomaba como referencia la calificación numérica que se encuentra en la página web de donde fueron extraídos los comentarios para su clasificación, por ejemplo, entre 1 y 3 se etiqueta como una reseña negativa y entre 4 y 5 como positiva.

En cuanto a los modelos o algoritmos de aprendizaje automático utilizados para evaluar el rendimiento del clasificador (ver **Error! Reference source not found.**), se destacan Naive Bayes, con presencia en 8 artículos; de estos, en 7 se realiza una comparación con otros modelos y obtiene el mejor desempeño en uno de ellos. Las máquinas de vectores de soporte (SVM) también son ampliamente utilizadas, apareciendo en 8 artículos y superando a otros modelos en tres de ellos.

Los bosques aleatorios son mencionados en 5 artículos, obteniendo el mejor desempeño en dos casos tras ser comparados con otros modelos en cuatro artículos. Los árboles de decisión se destacan en 4 artículos, logrando el mejor desempeño en uno de ellos. Por último, BERT, basado en la arquitectura de Transformers, es mencionado en 4 artículos, obteniendo el mejor rendimiento en uno de ellos después de una comparación con otros nueve modelos.

De igual manera todos los modelos y librerías se encuentran relacionados con cada artículo en el Apéndice A.

Figura 9. *Herramientas para análisis de sentimiento en los artículos de la revisión de literatura.*

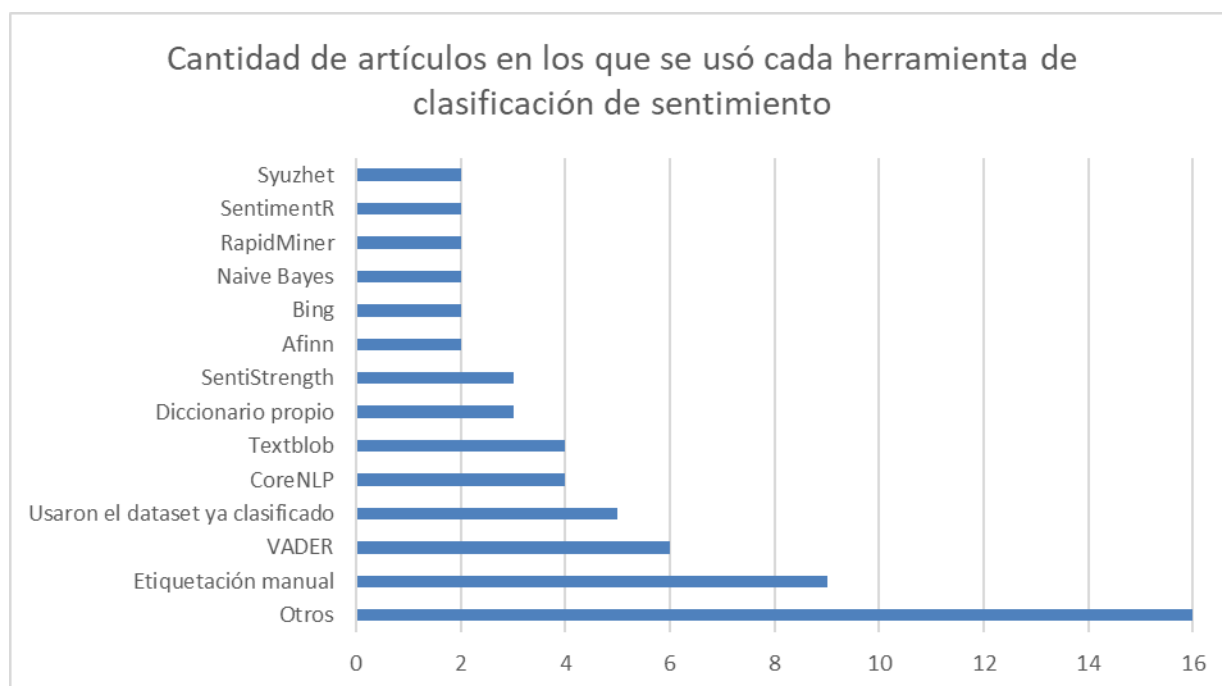
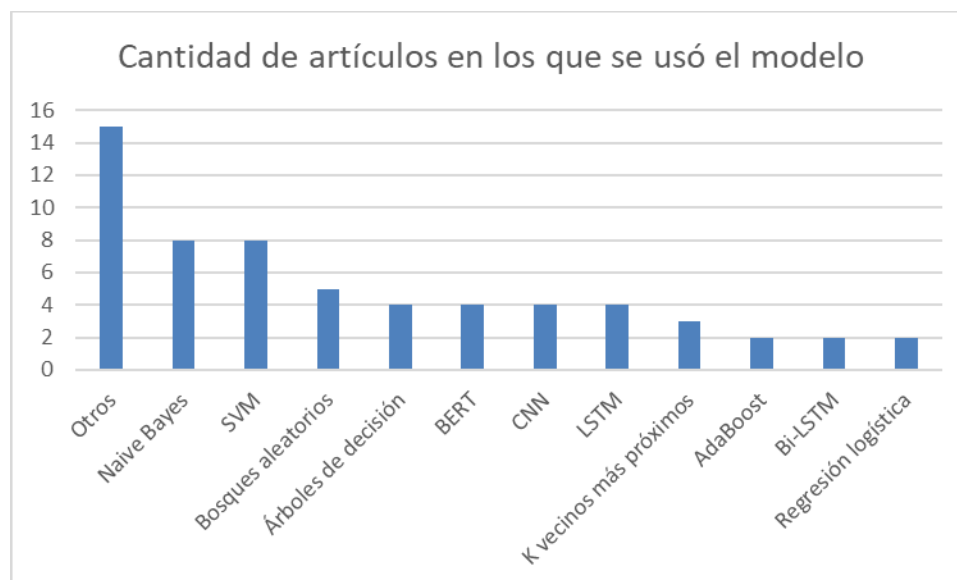


Figura 10. Modelos de aprendizaje automático en los artículos de la revisión de literatura.

Como principal limitación al problema de análisis de sentimiento se encuentran el tema de los diccionarios y la escasez de estudios en idiomas diferentes al inglés. La mayoría de los diccionarios no reconocen el sarcasmo, la doble negación, las jergas, o los modismos (Oliveira et al., 2022b), y están adaptados a un contexto específico por lo que el mismo término puede tener un significado diferente y afectar la precisión de la polaridad del sentimiento. Por esta razón incluso si el análisis de texto es una herramienta muy útil se puede considerar que necesita de supervisión humana para aspectos como el etiquetado, de manera que minimice la probabilidad de error en estos casos (W. Yu et al., 2023).

6. Proceso de descubrimiento del conocimiento en bases de datos

6.1 Extracción y selección de datos

El proceso de selección de datos comienza con la exploración de la plataforma TripAdvisor y los hoteles disponibles para el área de Bucaramanga. Se encuentran 156 establecimientos categorizados en seis divisiones, de la siguiente manera: B&Bs y posadas (59), Hoteles (59), Pensiones (16), Hostales (10), Alojamientos especiales (8) y Apartamentos (4).

Al revisar los establecimientos en cada uno de los grupos, se encuentra que las categorías de B&Bs y posadas, Hoteles, y Alojamientos especiales, estaban conformadas en su totalidad por hoteles, a diferencia de las categorías restantes, razón por la que se descartan debido que los demás tipos de alojamiento no son el objeto de estudio del proyecto. De esta manera se limita la revisión inicial a 126 hoteles.

Se procede a extraer manualmente la información sobre la cantidad total de comentarios para cada uno de los hoteles, descartando 49 establecimientos que no que no tenían comentarios registrados en la plataforma. En el Apéndice B se encuentra la cantidad de comentarios obtenidos de la revisión inicial de los 77 hoteles.

Una vez finalizado este proceso, es necesario limitar la información por fecha ya que algunos de los hoteles tienen registrado el último comentario varios años atrás, lo que indica que la información para estos hoteles no está actualizada y por lo tanto no podría realizarse la identificación de patrones en la calidad de servicio en todos los hoteles durante el mismo lapso. Para definir el rango de tiempo más apropiado para la extracción de datos, se toman como referencia las resoluciones y los decretos expedidos por el Gobierno Nacional en el contexto de la emergencia sanitaria del COVID 19.

En la Resolución 1285 de 2020 el Gobierno Nacional presentó las medidas generales de bioseguridad que debían adoptar los alojamientos en hoteles, apartahoteles, centros vacacionales y otros tipos, con el fin de mitigar, controlar y dar un manejo adecuado a la pandemia del COVID 19. Dentro de dichas medidas se encuentran la suspensión del uso de minibar; la limitación de ocupación en habitaciones, espacios comunes, áreas de gimnasio y piscinas; la adaptación de protocolos rutinarios para los huéspedes como la toma de temperatura y el diligenciamiento de formularios para el registro de las condiciones de salud; la priorización de ventilación natural en las habitaciones en lugar del uso de aire acondicionado; entre otras (Resolución 1285, 2020), por tanto estas medidas afectan la experiencia del turista, llegando a restringir la disponibilidad o incluso a impedir ciertos servicios que contribuyen a la satisfacción del cliente. Adicionalmente, el Decreto 1168 de 2020 establece el inicio del aislamiento selectivo y el distanciamiento individual responsable desde el primero de septiembre de 2020, con el que se dio fin al aislamiento obligatorio en Colombia y se comenzaron a tomar medidas para la reactivación del comercio en el país (DECRETO 1168, 2020) .

Para los 77 hoteles restantes se extrae inicialmente: la información de la categoría de estrellas, la cantidad de comentarios desde septiembre de 2020 hasta diciembre de 2023, el lugar que ocupan en la plataforma según la calificación Calidad/precio otorgado por TripAdvisor basada en la calificación de los usuarios, la disponibilidad, los precios, la popularidad de reservación y la ubicación del hotel (Equipo de administración Tripadvisor, 2019). Para asegurar que se consideraran los hoteles que tienen información completa en todos los periodos de tiempo, se descartan aquellos hoteles cuyo último comentario estuviera registrado antes de septiembre de 2020, teniendo en cuenta el contexto de la pandemia del COVID 19, el confinamiento y las

consecuencias para el sector turismo en términos de regulaciones y limitaciones por los protocolos implementados en esa época. Con la limitación por fechas se obtuvo un total de 42 hoteles.

Como se ilustra en la Figura 11 **Error! Reference source not found.** se utiliza el principio de Pareto para identificar qué establecimientos son los más influyentes según la cantidad de reseñas disponibles, el 20% de los hoteles contribuyen con más del 80% del total de comentarios, por lo que se seleccionan finalmente los seis hoteles de la **Error! Reference source not found.** para comenzar con el proceso de extracción de datos.

Figura 11. Cantidad de comentarios por hotel

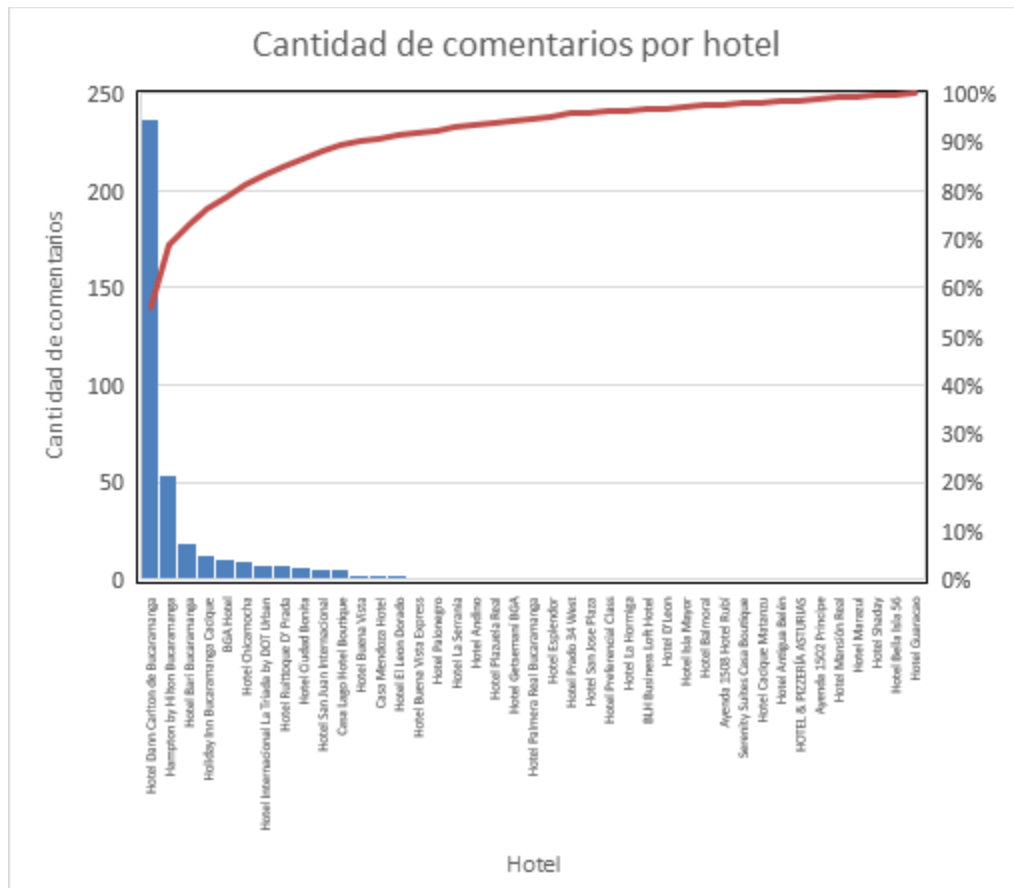


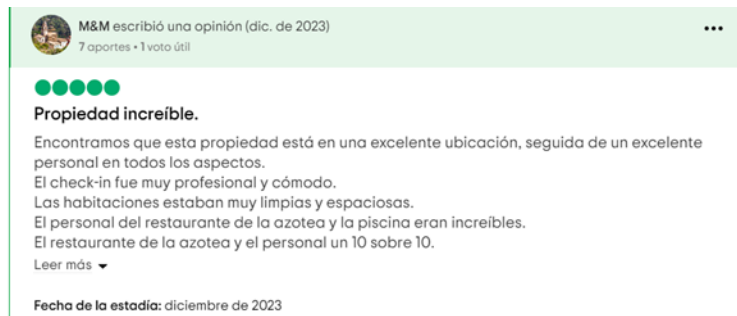
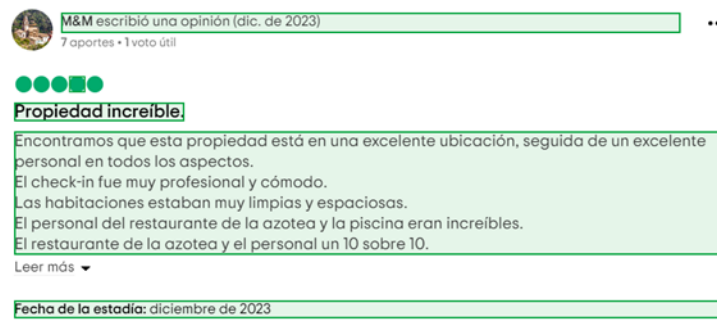
Tabla 2. Cantidad de comentarios por hotel

| NOMBRE DEL HOTEL | CANTIDAD DE COMENTARIOS |
|------------------|-------------------------|
|------------------|-------------------------|

| | |
|-----------------------------------|------------|
| Hotel Dann Carlton de Bucaramanga | 239 |
| Hampton by Hilton Bucaramanga | 55 |
| Hotel Barí Bucaramanga | 22 |
| Holiday Inn Bucaramanga Cacique | 13 |
| BGA Hotel | 11 |
| Hotel Chicamocha | 10 |
| TOTAL | 350 |

Se utiliza el 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 al usuario. Por medio del centro de ayuda del software se encuentra una guía para la extracción de reseñas de hoteles desde TripAdvisor.

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. Se comienza seleccionando los datos generales del hotel, en este caso son: el nombre del hotel, el lugar que ocupa en la ciudad según la clasificación de los usuarios en la plataforma, la cantidad de opiniones y las estrellas. Después, se seleccionan las dos primeras reseñas y seguidamente, de la primera reseña, se toman los datos requeridos. Esto dará al software la guía de la estructura del contenido a extraer de ahí en adelante, es decir, deberá tomar todos los elementos que tengan estructura de reseña como se ilustra en la **Error! Reference source not found.** y los contenidos específicos dentro de ella (ver **Error! Reference source not found.**).

Figura 12. Selección de la estructura de la reseña en Octoparse**Figura 13. Selección de los datos dentro de la reseña.**

A partir de esto, el software automáticamente extrae de la misma manera los datos de las demás reseñas de la página, teniendo en cuenta que se muestran cinco reseñas por página. A medida que se seleccionan los datos, se muestra en la parte inferior lo que la herramienta está extrayendo, tal como se muestra en la Figura 14.

Figura 14. Interfaz de trabajo de Octoparse

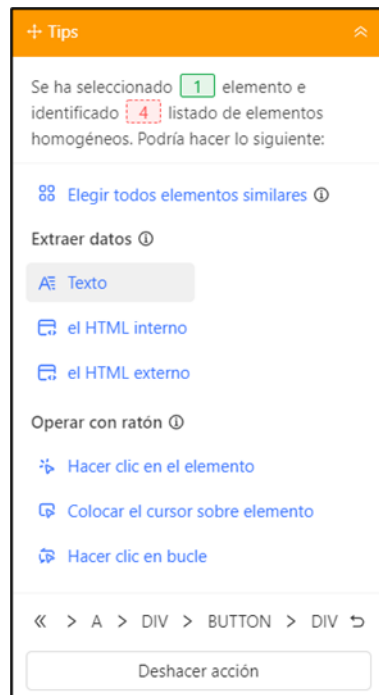
| Campos de datos | | Núm. | Fecha | Contenido | Tiempo | Titulo1 | Contenido1 | Camf + | Operaciones |
|-----------------|--|------|--------------------------|---------------------------|--------------------------|-------------------------|-------------------------|---------------|-------------|
| ▼ Página1 | | 1 | VICTORSORZA escri... | Hotel bonito, bien u... | Fecha de la estadia: ... | Hotel bonito pero p... | | | 📄 🗑️ |
| Extraer datos | | 2 | pedroluis1802 escrib... | Me sentí bien atendi... | | Llegada Segura | Me sentí bien atendi... | Fecha de la e | 📄 🗑️ |
| Extraer datos2 | | 3 | luisnoguera0712 esc... | Lugar agradable , tie... | | Agradable lugar | | | 📄 🗑️ |
| | | 4 | pmarquez2023 escri... | El servicio es excelen... | | EXCELENTE HOTEL | | | 📄 🗑️ |
| | | 5 | Iggiraldo escribió un... | Es un hotel bien ubic... | | Hotel Bien Ubicado y... | | | 📄 🗑️ |

En la visualización previa de los datos se pueden agregar nuevos campos con el fin de seleccionar los datos necesarios que no se evidencian en la misma columna porque no fueron leídos

por el software. Por ejemplo, en el caso de la Figura 14 en la columna Tiempo, solo se ve la fecha de estadía para el primer comentario, por lo que se agregó un campo en la columna 7 que extrae la fecha deseada del segundo comentario, esto sucede cuando la estructura de las reseñas no es igual en la primera página.

Adicionalmente, al seleccionar los elementos deseados el software muestra automáticamente un panel de Tips que sugiere las opciones a realizar según el tipo de elemento (Figura 15).

Figura 15. Panel de sugerencias de Octoparse



En este caso el campo 1 hace referencia al usuario y la fecha en la que escribe el comentario, el segundo al contenido del comentario, el tercero a la fecha de estadía, el cuarto al título de la reseña, los demás son campos complementarios que extraen la misma información cuando no pudo ser leída al seleccionarse. El flujo de trabajo se puede observar en la Figura 16.

Figura 16. Flujo de trabajo en Octoparse

The screenshot displays the Octoparse interface. On the left, the target website is shown, featuring hotel listings for Hampton by Hilton Bucaramanga and similar establishments like Hotel Dann Carlton de Bucaramanga, Hotel Chicamocha, and Holiday Inn Bucaramanga. On the right, a workflow diagram is visible, detailing the steps for data extraction, including navigating to the page, extracting data, handling pagination, clicking on elements, and clicking on specific data points. Below the workflow, a 'Vista previa de datos' (Data Preview) table is shown, containing the following data:

| Campos de datos | Núm. | Fecha | Contenido | Tiempo | Título1 | Contenido1 | Camj | Operaciones |
|-----------------|------|--------------------------|---------------------------|--------------------------|-------------------------|-------------------------|---------------|-------------|
| Extraer datos | 1 | VICTORSORZA escrib... | Hotel bonito, bien u... | Fecha de la estadía: ... | Hotel bonito pero p... | | | |
| Extraer datos2 | 2 | pedroius1802 escrib... | Me sentí bien atendi... | | Llegada Segura | Me sentí bien atendi... | Fecha de la e | |
| | 3 | luisnoguera0712 esc... | Lugar agradable, tie... | | Agradable lugar | | | |
| | 4 | pmarquez2023 escri... | El servicio es excelen... | | EXCELENTE HOTEL | | | |
| | 5 | lggiraldo escribió un... | Es un hotel bien ubic... | | Hotel Bien Ubicado y... | | | |

Al finalizar con la selección de los datos se ejecuta el flujo de trabajo y se procede a exportar el archivo con la información.

Este proceso se realiza para dos hoteles de la Tabla 2 el Hampton by Hilton y el Dann Carlton. Sin embargo, se extrae la totalidad de comentarios disponibles en TripAdvisor para cada hotel y se descartaron manualmente en Excel los anteriores a septiembre de 2020, filtrando por la columna Fecha, con esto se tiene la base de datos con la que se realizará el análisis de sentimiento.

Para los cuatro hoteles restantes, se realizó la extracción manual de los datos, considerando que la cantidad de comentarios seleccionados por la limitación de fechas era muy pequeña en comparación a la cantidad total de comentarios en la plataforma.

6.2 Etiquetado manual del conjunto de datos

Una vez finalizada la extracción de la base de datos desde Octoparse en formato xlsx, para cada una de las reseñas se obtienen 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. El proceso de etiquetado se realiza para el título y contenido de la reseña.

La etiquetación se realiza de manera manual, es decir, se revisa cada uno de los 350 comentarios para determinar qué sentimiento expresaba. Para la detección de polaridad de los comentarios, se opta por escoger tres categorías: positivo, negativo y neutro. El proceso se lleva a cabo de manera independiente para el título y para el contenido de los comentarios, con lo que realmente se etiquetaron 700 entradas de texto. Para definir a qué clasificación pertenece se revisa cada registro y se asigna la etiqueta correspondiente basada en lo que se expresara en el comentario. Si se resaltan aspectos buenos sobre el hotel se asignaba la etiqueta de “Positivo”; si se presenta a modo de queja o se notaba en general el disgusto en el comentario, se le asigna la etiqueta “Negativo”; y se considera “Neutro” en dos ocasiones, primero, si el comentario no expresa un sentimiento sino que da alguna declaración que podría o no contener información específica sobre el hotel, o si mencionan aspectos negativos y positivos en el mismo comentario de manera que se pudieran equilibrar y ninguno de los dos sentimientos predominara sobre el otro. Sin embargo, algunos comentarios no son tan fáciles de identificar, ya que, al presentar los aspectos positivos y negativos al tiempo, no se tiene claridad acerca de cuál prevalecía sobre el otro. Para este caso se etiquetan como neutros dependiendo de si los dos sentimientos se compensan ya fuera en cantidad (por ejemplo, si mencionaba tres cosas agradables y tres desagradables), o en proporción de lo que se considera más relevante al momento de evaluar el servicio de estadía. Para

mayor claridad, se ejemplifica en la **Error! Reference source not found.** la asignación de etiquetas para cada sentimiento y la dificultad que se presentó. Se incluyen solo contenidos de los comentarios ya que no se presentó mayor dificultad al clasificar los títulos, que son en su mayoría frases con menos de 10 palabras.

Tabla 3. *Etiquetación manual de los comentarios*

| COMENTARIO TEXTUAL EXTRAIDO DE TRIPADVISOR | ETIQUETA ASIGNADA | DIFICULTAD EN LA ETIQUETACIÓN | EXPLICACIÓN DE LA ETIQUETA |
|--|-------------------|-------------------------------|--|
| <p>“Encontramos que esta propiedad está en una excelente ubicación, seguida de un excelente personal en todos los sentidos. El check-in fue muy profesional y cómodo. Las habitaciones estaban muy limpias y espaciosas. El personal del restaurante de la azotea y la piscina eran increíbles. El restaurante de la azotea y el personal un 10 sobre 10.”</p> | Positivo | Fácil | Solo menciona aspectos positivos. |
| <p>“Me gusto mucho el lobby del hotel y los restaurantes. Los desayunos exquisitos. La piscina un poco pequeña. El despertador seguro quedo programado y nos despertó a la madrugada. La habitación confortable también nos gusto.”</p> | Positivo | Media | A pesar de mencionar aspectos que no son positivos, como el tamaño de la piscina y el incidente con el despertador, no los expresa a modo de queja o en un tono desfavorable, y en general nombra más aspectos positivos y que son más importantes al momento de evaluar un hotel (el estado de la habitación, el espacio de la entrada y el servicio de restaurante). |
| <p>“No hay roomservice las 24hrs mi vuelo se retraso llegue 11pm y no tenían ningún servicio de alimentación disponible. Los desayunos no hay variedad tome un sandwich y el jamón olía mal y pedí chocolate caliente y estaba frio”</p> | Negativo | Fácil | Solo menciona cosas que no fueron de su agrado. |
| <p>“La increíble propiedad y el personal de limpieza fueron brillantes. El personal de recepción estaba más preocupado por contestar los teléfonos móviles que por hacer que los huéspedes se registraran</p> | Negativo | Media | Comienza destacando el espacio y el personal de limpieza, sin embargo, se extiende mucho más en la parte negativa del personal de la recepción y finaliza utilizando la palabra “doloroso” para |

correctamente y los tiempos de espera para la comida en el bar/restaurante de la piscina eran dolorosos.”

describir el tiempo de espera, por lo cual la parte positiva del comentario no logra equilibrarse con lo negativo y predomina el sentimiento de disgusto.

“Hola, buenos días. La atención durante la estadía fue perfecta. Sin embargo en el posterior, he pedido se me envíen las facturas electrónicas de mis comidas en el restaurante y solo han enviado una. Podrían por favor ayudarme enviando la faltante? Que debo hacer para recibirlas?”

Neutro

Fácil

Menciona que su estadía fue perfecta, pero predomina la neutralidad, ya que se hace una pregunta directamente sin expresar ningún sentimiento o describir aspectos del hotel.

“Me encanto que realmente la piscina esté climatizada la única observación es q cuando el jacuzzi sea utilizado por el spa no lo pongan en servicio hasta q no esté completamente limpio fue lo único incómodo en mi estadía volveré al hotel encantada”

Neutro

Media

Pese a la incomodidad aclara que volvería con gusto, por lo que se compensan las cosas positivas con las negativas.

“Algo para mejorar es tener una muy buena fotocopiadora a blanco y negro y a color, pues la que tienen es de muy mala calidad, entonces uno tiene que salir a buscar una papelería alrededor del hotel y no es fácil encontrar una cerca.”

Neutro

Difícil

Al dar la queja por la fotocopiadora termina describiendo una situación que podría ponerlo en riesgo. Aunque el estado de la fotocopiadora no es algo usualmente relevante al momento de escoger un hotel, para esta persona lo fue y se extiende mucho en el tema. Sin embargo, termina resaltando que los demás aspectos fueron excelentes, por lo que se considera un comentario neutro.

“Exponiendose uno a robos pues en la zona se observa inseguridad para los transeuntes.

Pueden cobrar las fotocopias y uno con gusto las paga. Por lo demas todo muy bien.

La comida del restaurante excelente.

La habitacion excelente.”

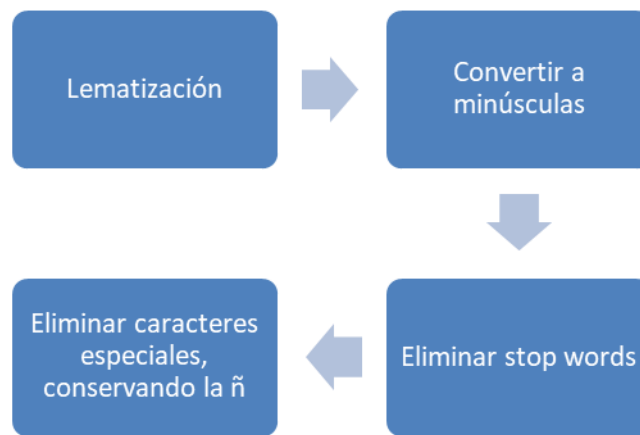
Al finalizar el proceso de etiquetado se tiene para los títulos, 242 positivos, 56 neutros y 52 negativos, y para los contenidos, 246 positivos, 19 neutros y 85 negativos. La diferencia notable entre la cantidad de neutros por cada categoría se debe a que en los títulos se encuentran registros como “Visita a Bucaramanga”, “Estadía”, “Gracias”, entre otros, que no expresan ningún sentimiento, en cambio los contenidos tienen más información para determinar la polaridad.

6.3 Preprocesamiento de los datos

El preprocesamiento es uno de los pasos más importantes en las tareas del Procesamiento de Lenguaje Natural (por sus siglas en inglés PLN) (Puh & Bagić Babac, 2023), como se describió en la **Error! Reference source not found.**, esta etapa permite revisar la calidad de los datos y aplicar procesos para mejorarla.

Con el conjunto total de datos etiquetado se realiza un análisis inicial por medio de la extracción de bigramas y las nubes de palabras por cada sentimiento. Para esto primero es necesario realizar el preprocesamiento.

Se opta por comenzar con una revisión de ortografía de manera manual con la herramienta de revisión de Microsoft, ya que al ser contenido generado por usuarios es muy probable que se encuentren errores de tipografía y por esta razón en el preprocesamiento las librerías utilizadas no reconozcan algunas palabras, generando imprecisión en la limpieza del texto. Al finalizar la revisión ortográfica, se utiliza el lenguaje de Python para programar y ejecutar el siguiente proceso en el entorno de Google Colab (Figura 17):

Figura 17. *Proceso de limpieza*

La estructura del proceso inicial de limpieza se define después de realizar varias pruebas entre los pasos ya definidos. En un principio se considera iniciar con la conversión a minúsculas, seguida por la eliminación de acentos y stopwords, aplicando por último el proceso de lematización. Se toma la decisión de comenzar con la lematización utilizando la librería Spacy, teniendo en cuenta que al realizar este proceso algunos nombres propios no son reconocidos al estar en minúscula y haber eliminado las tildes. Algunas palabras como “colombia”, “chicamocho”, “cacique”, entre otras, son transformadas por el código a una especie de verbo como “colombiar”, “chicamochaber”, “cacicar”. Adicionalmente, sustituye los verbos conjugados por el verbo en infinitivo y el pronombre, por ejemplo, “acostarme” era reemplazado por “acostar yo” y “despedirse” se convertía en “despedir él”, por esta razón después del proceso, el conjunto de palabras era mayor al conjunto inicial y los pronombres tenían frecuencias muy altas. Al comenzar con la lematización se asegura que los significados de muchas palabras no sean alterados y que los pronombres no aparezcan nuevamente después de haber sido eliminados por ser parte de las stopwords.

Después, el texto se cambia a minúsculas considerando que el diccionario de stopwords está definido en minúsculas e incluye tildes. Para la conversión se utiliza el método integrado de `Python.lower()`. Una vez todas las palabras están en minúsculas se eliminan las stopwords nombrando una lista predefinida del conjunto de librerías NLTK que incluye términos como “él”, “sí”, “con”, “de”, “la”, “los”, etc. Se examina el resultado y se encuentra que algunas palabras que no aportan información sobre aspectos de los hoteles tienen frecuencia muy alta dentro de los comentarios, lo cual no es útil al momento del análisis. Por esta razón, se añaden a la lista de stopwords términos como: “si”, “ser”, “mas”, “haber”, “ir”, “así”, “pm”, “am”, entre otros, para eliminarlos. Luego se eliminan las tildes y caracteres especiales con el módulo `re` de Python preservando la letra ñ para no alterar las palabras que la contienen. Adicionalmente, como último paso se define un diccionario de reemplazos con base en las palabras que la librería de Spacy no transforma de manera correcta, esto sucede sobre todo con sustantivos y adjetivos femeninos. Se logra la identificación de estas palabras al revisar las frecuencias y hacer una comparación con el archivo inicial para saber a qué palabras corresponden realmente aquellas que no tienen sentido. Dentro del diccionario se definen reemplazos como: "azotear" por "azotea", "barmar" por "barman", "call" por "calle", "ducho" por "ducha", "hospedir" por "hospedar", "humedade" por "humedad", entre otros. También se reemplaza “malo” por “mal” debido a que a pesar de que ambas palabras tienen sentido, las dos tienen frecuencias muy altas y representan lo mismo.

Para facilidad de manejo en Python se extraen el título y el contenido de los comentarios del dataframe. Estos se incluyen en listas que después son convertidas en cadenas para poder separar por palabras o tokens. Posteriormente, se extraen los grupos de dos palabras consecutivas llamados bigramas y se construyen las nubes de palabras por medio de los paquetes de `matplotlib` y `wordcloud`.

Con la clasificación de los comentarios se generan 6 dataframes, diferenciando cada sentimiento para los títulos y contenidos. Con esto se generan las nubes de palabras para cada dataframe con los bigramas más frecuentes en los comentarios negativos, neutros y positivos. Las nubes de palabras dan una representación visual que ayuda a la comparación de los dos grupos de hoteles, en la que entre mayor es la frecuencia de la palabra su tamaño es mayor en la nube.

En la **Error! Reference source not found.** se muestra que para los títulos resaltan términos como “mal experiencia”, “momento desagradable” y “estancia desagradable”, que claramente expresan sentimientos negativos sobre el hotel. Se evidencia que varios bigramas hacen referencia a la insatisfacción con la atención y el servicio (“mal atención”, “desmejorar atención”, “pesimo servicio”, “peor servicio”) y se identifican algunos aspectos adicionales como la deficiencia en las instalaciones y la experiencia en el parqueadero.

A diferencia de los títulos, los contenidos no tienen muchos bigramas que evidencien la insatisfacción con el hotel como se presenta en la Figura 18. Sin embargo, ofrecen aspectos más específicos que dan a entender, por ejemplo, que se presentan varios inconvenientes con el aire acondicionado como “aire acondicionado” y “habitación aire”, con la comida, el *check in*, la seguridad del hotel y el personal de recepción.

Figura 18. Nube de palabras de bigramas negativos para Título.



Figura 19. Nube de palabras de bigramas negativos para Contenido.



Dentro de los bigramas de los títulos clasificados como neutros (Figura 20), se encuentran algunos términos como “experiencia Dann”, “visita Bucaramanga”, “fin semana”, “primera vez”, que no proporcionan información sobre el sentimiento de los usuarios. Aunque en su mayoría la Figura 21 contiene bigramas similares a los títulos como “booking.com”, “permanecer así”, “comida restaurante”, entre otros que expresan ninguna opinión, también se encuentran términos como “buena comida”, “buena atención” y “buen día”, que expresan algo positivo. Esto sucede porque, como se trató en **Error! Reference source not found.**, 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 20. Nube de palabras de bigramas neutros para Título



Figura 21. Nube de palabras de bigramas neutros para Contenido

Para los bigramas de los títulos positivos resaltan las palabras “buen” y “excelente” haciendo referencia a aspectos como la ubicación, el servicio, la atención, el personal y el precio (**Error! Reference source not found.**). Los términos encontrados en los contenidos no varían mucho con respecto a los títulos, pero dan un poco más de información resaltando características como la comodidad: “habitación cómoda”, “habitación amplia”, “habitación confortable”; la comida: “desayuno variado”, “comida excelente”, “buena comida”, “comida deliciosa”, “restaurante buen”; el aseo: “habitación limpia”, “limpio personal”, “habitación impecable”.

6.4 Etiquetado de los datos con modelos pre entrenados

En este paso se utilizan las librerías de Pysentimiento y VADER para clasificar los comentarios y posteriormente comparar su desempeño teniendo como base la etiquetación manual.

Pysentimiento es un kit de herramientas de Python para tareas de NPL. Su estructura está basada en la arquitectura de BERT, y fue entrenada en diferentes tareas: análisis de sentimiento, detección de emociones y discurso de odio. La librería funciona para texto en español, italiano, inglés y portugués. Para el análisis de sentimiento toma las tres polaridades, positivo, negativo y neutro, e indica la probabilidad de que el texto pertenezca a cada una de ellas (Pérez et al., 2021).

Por su parte, VADER (Valence Aware Dictionary and sEntiment Reasoner), es un módulo de NLTK de Python que fue entrenado con tweets y funciona como una librería multilingüaje. Utiliza un enfoque basado en léxico para determinar el sentimiento de un texto, es decir, se basa en un conjunto predefinido de palabras y reglas para asignar una puntuación de sentimiento a cada palabra en el texto. Estas puntuaciones se combinan para calcular un puntaje compuesto de sentimiento para el texto de entrada (Hutto & Gilbert, 2014).

Se realiza la detección de polaridad para los títulos y los contenidos. En la **Error! Reference source not found.** se observa el resultado del análisis con Pysentimiento, en el que da la polaridad del comentario (NEG: negativo, NEU: neutro, POS: positivo) y la probabilidad de cada una de las polaridades comenzando por la mayor de ellas.

Figura 24. Fragmento del archivo de Excel con la clasificación de los comentarios por Pysentimiento.

| Título | Contenido | sentiment Título | sentiment Contenido |
|------------|--------------------------------|---|---|
| nunca dec | cadena hotel gustar llegar ult | AnalyzerOutput(output=NEU, probas={NEU: 0.516, NEG: 0.359, POS: 0.125}) | AnalyzerOutput(output=NEU, probas={NEU: 0.500, POS: 0.490, NEG: 0.010}) |
| excelente | hotel dann carlton bucarame | AnalyzerOutput(output=POS, probas={POS: 0.930, NEU: 0.066, NEG: 0.004}) | AnalyzerOutput(output=POS, probas={POS: 0.881, NEU: 0.113, NEG: 0.005}) |
| descanso | excelente instalacion buen a | AnalyzerOutput(output=NEU, probas={NEU: 0.489, NEG: 0.365, POS: 0.147}) | AnalyzerOutput(output=POS, probas={POS: 0.807, NEU: 0.181, NEG: 0.013}) |
| fabuloso h | excelente hotel servicio habi | AnalyzerOutput(output=POS, probas={POS: 0.941, NEU: 0.053, NEG: 0.006}) | AnalyzerOutput(output=POS, probas={POS: 0.945, NEU: 0.050, NEG: 0.005}) |
| excelente | extraordinario hotel buen ub | AnalyzerOutput(output=POS, probas={POS: 0.821, NEU: 0.165, NEG: 0.014}) | AnalyzerOutput(output=POS, probas={POS: 0.963, NEU: 0.035, NEG: 0.003}) |
| buen hotel | buen hotel buena atencion s | AnalyzerOutput(output=NEU, probas={NEU: 0.497, POS: 0.368, NEG: 0.136}) | AnalyzerOutput(output=POS, probas={POS: 0.560, NEU: 0.425, NEG: 0.015}) |
| cada vez n | cada vez mejor gracias pode | AnalyzerOutput(output=POS, probas={POS: 0.832, NEU: 0.150, NEG: 0.018}) | AnalyzerOutput(output=POS, probas={POS: 0.961, NEU: 0.037, NEG: 0.002}) |
| excelente | excelente servicio personal f | AnalyzerOutput(output=POS, probas={POS: 0.917, NEU: 0.074, NEG: 0.009}) | AnalyzerOutput(output=POS, probas={POS: 0.931, NEU: 0.063, NEG: 0.006}) |
| fin semana | gustar lobby hotel restaurant | AnalyzerOutput(output=POS, probas={POS: 0.521, NEU: 0.409, NEG: 0.071}) | AnalyzerOutput(output=POS, probas={POS: 0.638, NEU: 0.352, NEG: 0.010}) |
| excelente | excelente ubicacion servicio | AnalyzerOutput(output=NEU, probas={NEU: 0.478, POS: 0.319, NEG: 0.203}) | AnalyzerOutput(output=POS, probas={POS: 0.914, NEU: 0.078, NEG: 0.008}) |

La salida de VADER (**Error! Reference source not found.**) es similar a la de Pysentimiento, da la probabilidad de cada polaridad (neg: negativo, neu: neutro, pos: positivo). Sin embargo, difiere en que proporciona las tres en el mismo orden sin importar cuál fue la mayor, y en que puede asignarle la probabilidad de 1.0 a alguna de ellas.

Figura 25. Fragmento del archivo de Excel con la clasificación de los comentarios por VADER.

| Título | Contenido | sentiment Título | sentiment Contenido |
|-------------|-------------|--|--|
| Deficiencia | Las instala | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0} | {'neg': 0.13, 'neu': 0.838, 'pos': 0.032, 'compound': -0.7863} |
| Propiedad | Encontrar | {'neg': 0.0, 'neu': 0.208, 'pos': 0.792, 'compound': 0.5859} | {'neg': 0.0, 'neu': 0.708, 'pos': 0.292, 'compound': 0.9593} |
| Lindo hote | Súper limp | {'neg': 0.0, 'neu': 0.204, 'pos': 0.796, 'compound': 0.5994} | {'neg': 0.022, 'neu': 0.557, 'pos': 0.421, 'compound': 0.9856} |
| nunca me | esta cader | {'neg': 0.0, 'neu': 0.478, 'pos': 0.522, 'compound': 0.2924} | {'neg': 0.0, 'neu': 0.821, 'pos': 0.179, 'compound': 0.9593} |
| Excelente | El Hotel D | {'neg': 0.0, 'neu': 0.575, 'pos': 0.425, 'compound': 0.5719} | {'neg': 0.0, 'neu': 0.698, 'pos': 0.302, 'compound': 0.9612} |
| Descanso | Excelentes | {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0} | {'neg': 0.0, 'neu': 0.697, 'pos': 0.303, 'compound': 0.9298} |
| Buenísimo | Excelente | {'neg': 0.0, 'neu': 0.328, 'pos': 0.672, 'compound': 0.6249} | {'neg': 0.0, 'neu': 0.626, 'pos': 0.374, 'compound': 0.9747} |
| La mejor r | He viajado | {'neg': 0.0, 'neu': 0.602, 'pos': 0.398, 'compound': 0.765} | {'neg': 0.024, 'neu': 0.562, 'pos': 0.414, 'compound': 0.9904} |
| Fabuloso f | Excelente | {'neg': 0.0, 'neu': 0.469, 'pos': 0.531, 'compound': 0.5267} | {'neg': 0.0, 'neu': 0.539, 'pos': 0.461, 'compound': 0.9792} |
| Excelente | Extraordin | {'neg': 0.0, 'neu': 0.316, 'pos': 0.684, 'compound': 0.7579} | {'neg': 0.0, 'neu': 0.701, 'pos': 0.299, 'compound': 0.9284} |

Para la comparación de resultados se emplea la biblioteca de aprendizaje automático *skit-learn* de Python, que permite obtener las matrices de confusión y las métricas necesarias para la revisión.

La matriz de confusión es una representación visual del desempeño de la predicción de un modelo en cuestión de aciertos y errores (Zapata, 2023). En este caso, es una matriz de clases múltiples y presenta una comparación de los comentarios que fueron clasificados correcta y erróneamente por las librerías en contraste con la etiquetación manual, que es el punto de

referencia. Para la explicación del funcionamiento de la matriz de confusión se usa como referencia Ardila, 2022, que aborda un problema similar. Los resultados de la diagonal de la matriz indican la cantidad de comentarios que fueron clasificados correctamente para cada polaridad, las demás celdas indican las que se clasificaron de manera incorrecta. A partir de los resultados se obtienen las siguientes métricas:

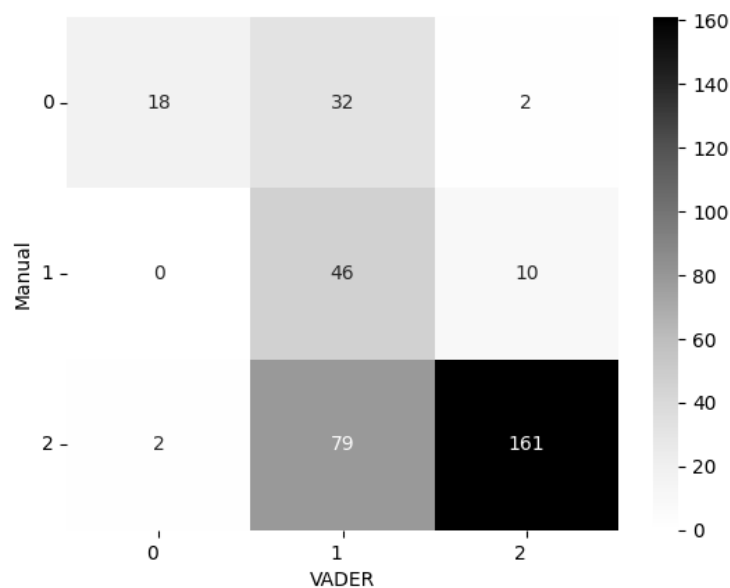
- **Precisión:** razón entre la cantidad de comentarios que se clasificaron correctamente en una categoría y la cantidad total (correctos e incorrectos) de comentarios que se clasificaron en esa categoría. Varía entre 0 y 1, donde 1 es el mejor rendimiento posible.
- **Recall:** razón entre la cantidad de comentarios que se clasificaron correctamente en una categoría y la cantidad de comentarios que en realidad pertenecían a esa categoría. Varía entre 0 y 1, donde 1 es el mejor rendimiento posible.
- **F1-score:** utiliza la precisión y el recall para calcular una métrica representativa. Varía entre 0 y 1, donde 1 es el mejor rendimiento posible.
- **Exactitud:** razón entre la cantidad de comentarios clasificados correctamente y el total de comentarios. Varía entre 0 y 1, donde 1 es el mejor rendimiento posible.
- **Coeficiente de correlación de Mathews (MCC):** utiliza todos los elementos de la matriz de confusión para dar un resultado imparcial del rendimiento, independientemente del número de elementos en cada categoría. Varía entre -1 y 1, donde 1 representa una predicción perfecta, 0 una predicción promedio y -1 una predicción inversa.

A continuación, se presentan los resultados de las dos librerías comparadas con la etiquetación manual por medio de matrices de confusión y las respectivas métricas. En las matrices

de confusión se representan las tres categorías de manera numérica, siendo 0 el sentimiento negativo, 1 el neutro y 2 el positivo.

6.4.1 VADER

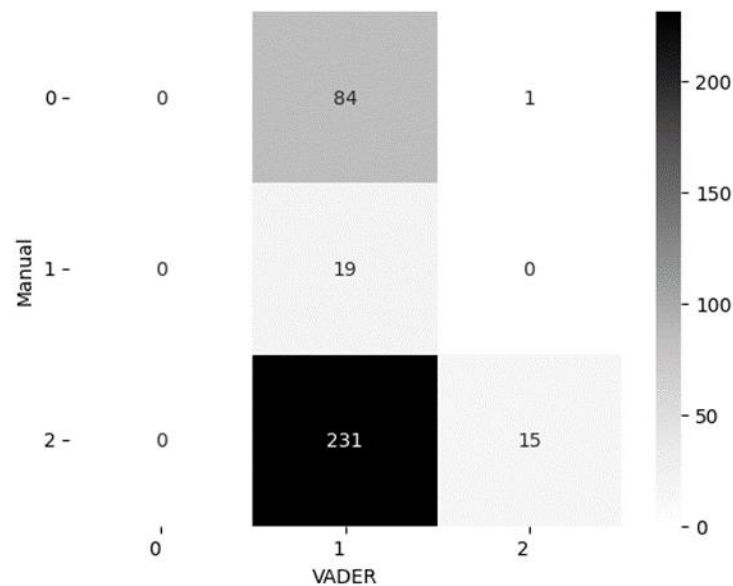
En la **Error! Reference source not found.** se presenta la matriz de confusión para los títulos identificados por VADER, mientras que en la **Error! Reference source not found.** se presentan las métricas obtenidas. Los comentarios neutros obtuvieron el menor valor de precisión y el mayor de recall. Esto sugiere que VADER clasificó correctamente pocos comentarios neutros en comparación a los que clasificó en total como neutros. En otras palabras, hubo muchos comentarios negativos y positivos que clasificó erróneamente como neutros (precisión), y clasificó muchos comentarios neutros correctamente en relación con los que debió haber clasificado como tales (recall). Es importante resaltar que, en este caso, la medida de recall alta no quiere decir necesariamente que el clasificador hace un buen trabajo, sino que, al estar identificando gran parte del total de comentarios bajo la categoría de neutros, también identificaría muchos correctamente en esa categoría.

Figura 26. Matriz de confusión del etiquetado manual VS VADER para los títulos.

Por otro lado, para los comentarios negativos, ocurre lo contrario: una precisión alta con el recall más bajo. Esto significa que VADER clasificó la mayoría correctamente en comparación con todos los que clasificó en esa categoría, pero identificó muy pocos de manera acertada en relación con los que clasificó erróneamente como neutros y positivos. Esto indica una dificultad para reconocer los comentarios negativos.

Finalmente, los comentarios positivos tuvieron la precisión y el F1-score más altos, lo que indica que en general fueron los comentarios que mejor clasificó.

Los resultados para los contenidos se observan en la Figura 27 y la Tabla 5. 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 27. Matriz de confusión de etiquetado manual VS VADER para los contenidos.

El recall es perfecto para los comentarios neutros, ya que clasificó correctamente en su totalidad los que debió haber clasificado como tales. La precisión es muy baja, porque, al igual que en los títulos, al clasificar la mayoría de los comentarios bajo esta categoría, identifica muchos negativos y positivos como neutros.

Para los comentarios positivos, la precisión es alta porque se identifican muy pocos como positivos, entonces la mayoría de estos eran en efecto positivos según la etiquetación manual. En cambio, para el recall, fueron muy pocos los que se clasifican como positivos en comparación a los que son realmente positivos. El MCC en este escenario es muy pequeño, 0.07, acorde con el desempeño del clasificador.

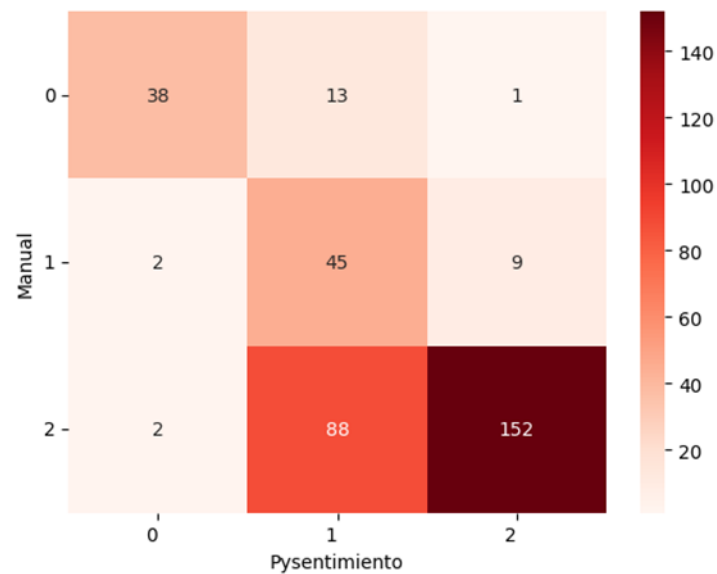
6.4.2 Pysentimiento

Con Pysentimiento se realizan las mismas pruebas obteniendo como resultado las matrices de las figuras **Error! Reference source not found.** (títulos) y **Error! Reference source not found.** (contenidos).

Respecto a los títulos, comenzando por los comentarios negativos, se encuentran valores aceptables de precisión, recall y F1-score (ver **Error! Reference source not found.**). La precisión es bastante alta ya que clasificó la mayoría de los negativos correctamente en comparación a la totalidad de los neutros y positivos que clasificó como negativos. El recall, junto con la matriz de confusión, indica que también clasificó la mayoría de los negativos acertadamente en comparación con los que eran verdaderos negativos, con lo que resulta en un F1-score alto.

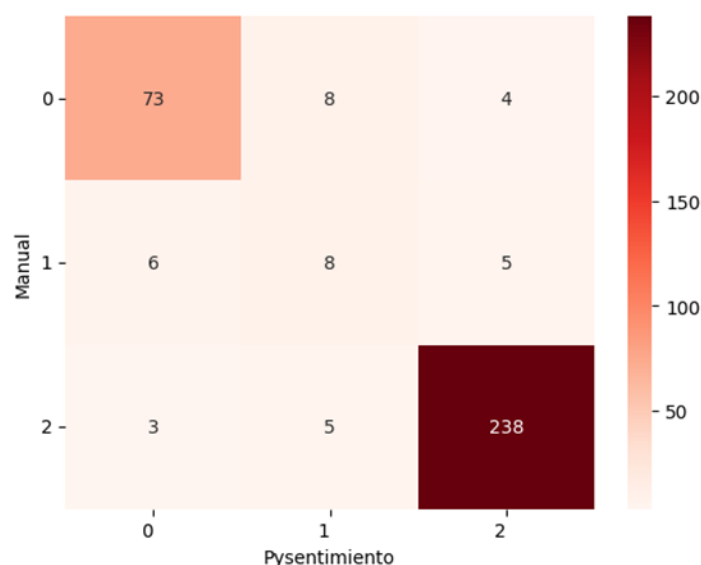
Para los comentarios neutros, la precisión es baja, evidenciando que clasifican muchos más comentarios erróneamente como positivos a los que clasifican acertadamente como neutros. El recall, por su parte, es mucho mayor incluso que el de los comentarios negativos, esto se debe a que clasifica como neutros la mayoría de los neutros verdaderos y, al mismo tiempo, clasifica muy pocos de ellos como negativos o positivos. Sin embargo, el F1-score no es alto, probablemente por su clasificación errónea de gran parte de los positivos verdaderos como neutros.

Finalmente, los comentarios positivos obtuvieron la precisión más alta ya que Pysentimiento identificó la mayoría correctamente y una proporción muy pequeña como negativos y neutros. El recall fue el más bajo de las tres categorías porque, a pesar de haber identificado la mayoría correctamente, clasificó 88 comentarios como neutros que son positivos según la etiquetación manual.

Figura 28. Matriz de confusión de etiquetado manual VS Pysentimiento para los títulos.

Para los contenidos, se obtuvieron medidas altas de precisión, recall y F1-score para los comentarios negativos y positivos. Sin embargo, en los comentarios neutros, el desempeño de las métricas es más bajo (Tabla 5). En este caso, clasificó la mayoría de los negativos correctamente en relación con el total clasificados como negativos y en relación a los verdaderos negativos; lo mismo ocurrió con los comentarios positivos.

Para los neutros, se clasificaron correctamente 8 comentarios como neutros, pero también clasificó 8 incorrectamente como negativos (precisión), aunque se clasificó la mayoría correctamente en relación con los neutros verdaderos (recall). El desempeño menor en las tres métricas para los comentarios neutros probablemente se deba a que tan solo 19 fueron etiquetados manualmente bajo esta categoría.

Figura 29. Matriz de confusión de etiquetado manual VS Pysentimiento para los contenidos.

Tras la revisión de las métricas es posible concluir que, en los dos escenarios (títulos y contenidos), el MCC de Pysentimiento es mayor al de VADER. Adicionalmente, se observa que Pysentimiento tuvo mejor desempeño para los contenidos que para los títulos, al contrario que VADER. Esto posiblemente se deba a que Pysentimiento pueda identificar el sentimiento con mayor precisión cuando cuenta con más información, recordando que los títulos tienen 10 palabras o menos.

En el caso de VADER, se observa que clasifican mejor los títulos en comparación con los contenidos. Cabe resaltar que, con la etiquetación manual, resultan más títulos neutros que contenidos neutros, por lo que esto no significa necesariamente que la librería funcione mejor para palabras o frases cortas en el caso de los títulos. Más bien, indica que en general tiende a clasificar los comentarios como neutros. Del mismo modo, había más contenidos negativos que títulos negativos en la etiquetación manual, categoría que a VADER le resultó más difícil reconocer.

Con esto se evidencia la importancia de la matriz de confusión, ya que permite tener una visión general, acertada y rápida de todas las categorías del problema de clasificación.

Tabla 4. Métricas de los clasificadores para los títulos.

| 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 5. Métricas de los clasificadores para los contenidos

| 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 | | |

6.4.1 Evaluación de la detección de polaridad con Pysentimiento

A partir de los resultados, se define que Pysentimiento funciona mejor como clasificador y se proponen realizar la clasificación llevando a cabo el preprocesamiento definido en **Error! Reference source not found.** de este documento y sus posibles combinaciones conservando la secuencia para verificar si pudiesen ayudar a la librería mejorar la detección de polaridad.

En la Tabla 6 se presentan las combinaciones de los pasos del preprocesamiento con la función asociada en el código de Python y un ejemplo del efecto que tiene individualmente en un comentario. Para este experimento no se incluye la conversión del texto a minúsculas, ya que Pysentimiento no distingue entre mayúsculas y minúsculas para la detección de polaridad¹

Tabla 6. *Pasos del preprocesamiento de datos*

| Paso del preprocesamiento | Función | Comentario original | Comentario procesado |
|----------------------------------|----------------------------|---|--|
| Corrección mecanográfica | Revisión manual | Las instalaciones en apariencia se amormizan con la categoría *****, pero resulta insólito que el aire acondicionado de la habitación que nos correspondió funcionara deficiente, a tal punto que la diferencia con la temperatura ambiente entre 26 a 28o c no era diferente. En el último reclamo el técnico, sin ningún empacho, dijo que no se podía bajar a menos de 22oC. Tenaz | Las instalaciones en apariencia se armonizan con la categoría *****, pero resulta insólito que el aire acondicionado de la habitación que nos correspondió funcionara deficiente, a tal punto que la diferencia con la temperatura ambiente entre 26 a 28o c no era diferente. En el último reclamo el técnico, sin ningún empacho, dijo que no se podía bajar a menos de 22oC. Tenaz |
| Lematización | word.lemma() | o de la habitación que nos correspondió funcionara deficiente, a tal punto que la diferencia con la temperatura ambiente entre 26 a 28o c no era diferente. En el ultimo reclamo el técnico, sin ninguno empacho , decir que no él poder bajar a menos de 22oc. Tenaz | el instalación en apariencia él amormizar con el categoría * * * * * , pero resultar insólito que el aire acondicionado de el habitación que yo corresponder funcionar deficiente , a tal punto que el diferencia con el temperatura ambiente entre 26 a 28o c no ser diferente . en el ultimo reclamo el técnico , sin ninguno empacho , decir que no él poder bajar a menos de 22oc. Tenaz |
| Eliminar stop words | stopwords.words("spanish") | entre 26 a 28o c no era diferente. En el ultimo reclamo el técnico, sin ningún empacho, dijo | instalaciones apariencia amormizan categoría resulta insólito aire acondicionado habitación correspondió funcionara deficiente tal punto diferencia temperatura ambiente diferente ultimo reclamo técnico ningún empacho dijo podía bajar menos oc tenaz |

¹ En el caso de Preprocesamiento de los datos, era necesario para que todos los bigramas con dichas palabras se incluyeran al momento de obtener la frecuencia y las nubes de palabras, por ejemplo, “habitacion Grande” habría sido diferente de “habitacion grande” si se hubiera obviado este paso.

| | | | |
|-----------------|-------------|--|---|
| Eliminar tildes | unidecode() | que no se podía bajar a menos de 22oC. Tenaz | Las instalaciones en apariencia se amormizan con la categoria *****, pero resulta insolito que el aire acondicionado de la habitacion que nos correspondio funcionara deficiente, a tal punto que la diferencia con la temperatura ambiente entre 26 a 28o c no era diferente. En el ultimo reclamo el tecnico, sin ningun empacho, dijo que no se podia bajar a menos de 22oC. Tenaz |
|-----------------|-------------|--|---|

Con las posibles combinaciones de los pasos de la tabla anterior, se obtienen los resultados de la Tabla 7, en donde se plantean un total de 16 escenarios para el título y contenido y se calcula el coeficiente de correlación de Matthews para contrastar los resultados.

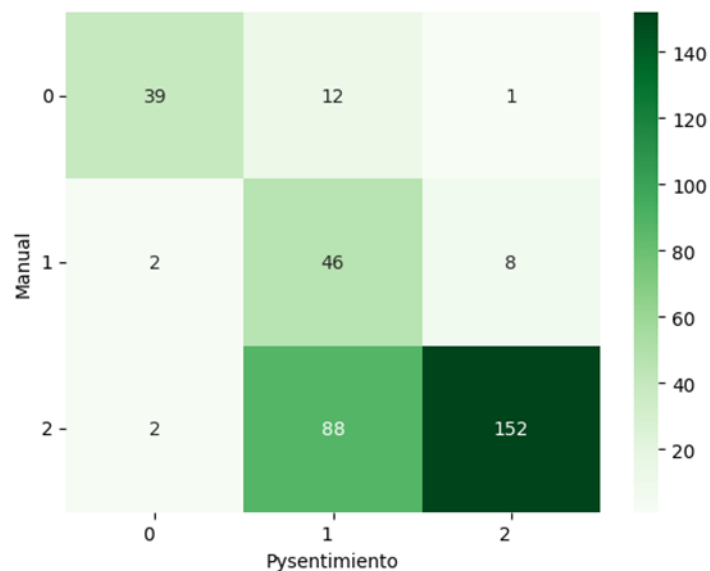
Tabla 7. Resultados del coeficiente de correlación de Matthews de la clasificación de Pysentimiento con las posibles combinaciones de preprocesamiento de datos.

| Escenario | Esquema del conjunto de datos | MCC | |
|-----------|---|---------------|---------------|
| | | Título | Contenido |
| 1 | Sin preprocesamiento | 0.5013 | 0.8005 |
| 2 | Corrección mecanográfica | 0.4961 | 0.8129 |
| 3 | word.lemma() + stop.words("spanish") + unidecode () | 0.405 | 0.5417 |
| 4 | word.lemma() + stop.words("spanish") | 0.3888 | 0.5463 |
| 5 | word.lemma() + unidecode() | 0.4297 | 0.7084 |
| 6 | stop.words("spanish") + unidecode() | 0.4139 | 0.6285 |
| 7 | word.lemma() | 0.418 | 0.7075 |
| 8 | stop.words("spanish") | 0.3892 | 0.6044 |
| 9 | unidecode() | 0.5141 | 0.7681 |
| 10 | Corrección mecanográfica + word.lemma() + stop.words("spanish") + unidecode() | 0.4029 | 0.5576 |
| 11 | Corrección mecanográfica + word.lemma() + stop.words("spanish") | 0.3888 | 0.556 |
| 12 | Corrección mecanográfica + word.lemma() + unidecode() | 0.4236 | 0.7202 |
| 13 | Corrección mecanográfica + stop.words("spanish") + unidecode() | 0.4117 | 0.6285 |

| | | | |
|-----------|---|--------|--------|
| 14 | Corrección mecanográfica + word.lemma() | 0.4142 | 0.7145 |
| 15 | Corrección mecanográfica + stop.words("spanish") | 0.3848 | 0.6005 |
| 16 | Corrección mecanográfica + unidecode() | 0.5115 | 0.761 |

Comenzando por los títulos, se encuentra que la librería tiene la clasificación más acertada al eliminar los caracteres especiales (Escenario 9 de la tabla 6). De los comentarios negativos, se puede decir que tiene un puntaje alto para las tres métricas (Tabla 7), Pysentimiento clasifica la mayoría de los negativos correctamente en comparación con los totales clasificados como negativos y con los clasificados como negativos manualmente. Con los comentarios neutros, la precisión fue baja porque clasifica más comentarios neutros incorrectamente como positivos, y el recall fue alto porque clasifica la mayoría correctamente con relación a la referencia de la etiqueta manual. Mientras que, para los positivos es el mismo caso de los negativos, aunque su precisión es un poco mayor porque hay muy pocos falsos positivos, y el recall es algo menor porque gran cantidad de los positivos en la referencia son clasificados como neutros por Pysentimiento (ver Figura 30).

Figura 30. Matriz de confusión de etiquetado manual VS Pysentimiento para los títulos después de eliminar caracteres especiales en el conjunto de datos.



Para los contenidos, se obtiene el MCC más alto realizando la corrección mecanográfica al conjunto de datos (Escenario 2 de la tabla 6). Al igual que en el caso de los títulos, se obtienen métricas altas para los comentarios negativos y positivos (Tabla 8). Esto se debe a que ambas categorías se identifican correctamente en relación con el total clasificado en cada categoría y con la referencia de etiquetación manual correspondiente. En cuanto a los neutros, el comportamiento es similar, pero debido a que la cantidad de comentarios en esta categoría es pequeña, hay muchos falsos neutros en proporción al total. Además, varios neutros son clasificados incorrectamente como negativos y positivos (ver Figura 31).

Figura 31. Matriz de confusión de etiquetado manual VS Pysentimiento para los contenidos después de realizar la corrección mecanográfica al conjunto de datos.

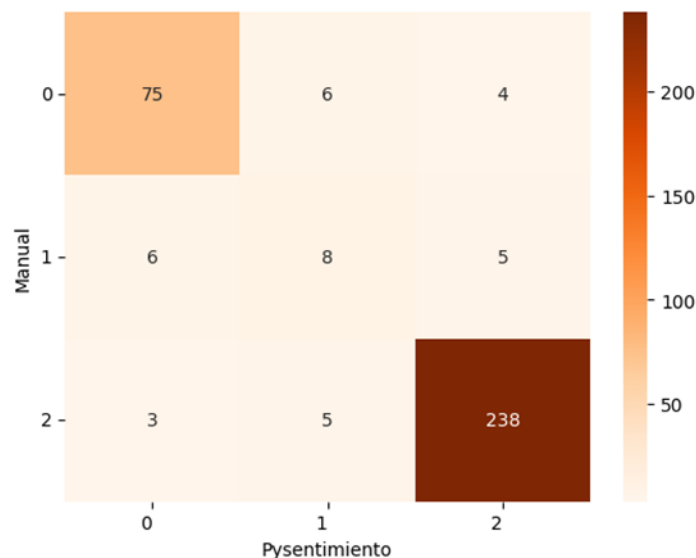


Tabla 8. Métricas de Pysentimiento para la clasificación de comentarios

| Manual VS Pysentimiento | | | | | | |
|-------------------------|-----------|--------|----------|---------|----------|-------|
| Títulos | Precisión | Recall | F1-score | Support | Accuracy | MCC |
| Negativo | 0.91 | 0.75 | 0.82 | 52 | 0.68 | 0.514 |
| Neutro | 0.32 | 0.82 | 0.46 | 56 | | |
| Positivo | 0.94 | 0.63 | 0.75 | 242 | | |
| Contenidos | Precisión | Recall | F1-score | Support | Accuracy | MCC |
| Negativo | 0.89 | 0.88 | 0.89 | 85 | 0.92 | 0.813 |
| Neutro | 0.42 | 0.42 | 0.42 | 19 | | |
| Positivo | 0.96 | 0.97 | 0.97 | 246 | | |

6.5 Análisis de sentimiento con BERT

En esta sección se busca entrenar el modelo BERT para la predicción en el análisis de sentimiento, y comparar su desempeño con los modelos utilizados para definir la configuración del modelo de Pysentimiento. Se tomó como referencia un código del repositorio de Github,

llamado Proyecto MAD, el cual está enfocado al análisis de sentimiento multiclase en español para tweets relacionados con la Educación superior en Colombia (Espinosa et al., 2023).

El documento de entrada para el código es la etiquetación manual de los comentarios. Se procede a asignarles una etiqueta numérica: 0 para negativo, 1 para neutro y 2 para positivo. El 80% de los comentarios se definen como el conjunto de entrenamiento el 20% restante como el conjunto de prueba.

Para la ejecución del modelo se utiliza la librería de scikit-learn y la función `ParameterGrid` para explorar las posibles combinaciones de hiper parámetros tomando como referencia la **Error! Reference source not found.** de (Pérez et al., 2021).

Después de definir la red de parámetros, se toma el texto original y se *tokeniza* para que pueda ser reconocido por la estructura que maneja BERT. También se define la variable de *padding lenght* para que se agreguen tokens nulos a los comentarios que tengan una longitud menor a la máxima. En este paso, se crea una clase que lea aleatoriamente el conjunto de datos y se crean las representaciones numéricas de cada token del conjunto de texto.

Más adelante, se crea el enmascaramiento, que devuelve un 1 para los tokens con contenido y 0 para los vacíos. Luego, se generan los embeddings, que convierten el texto en vectores numéricos para que puedan ser procesados, y se carga el modelo *bert-base-multilingual-cased*, que reconoce varios idiomas y distingue entre mayúsculas y minúsculas.

Con esto, se define la función para iterar según las distintas combinaciones de parámetros. Primero, se crean los conjuntos de entrenamiento y prueba. En segundo lugar, se define el modelo de la red neuronal, utilizando Pytorch. En este punto se carga el modelo pre entrenado de BERT, se define la variable de Dropout para evitar el sobreajuste, se define la red neuronal con una capa lineal. Después se define el optimizador utilizando AdamW, y finalmente se crea el loop para las

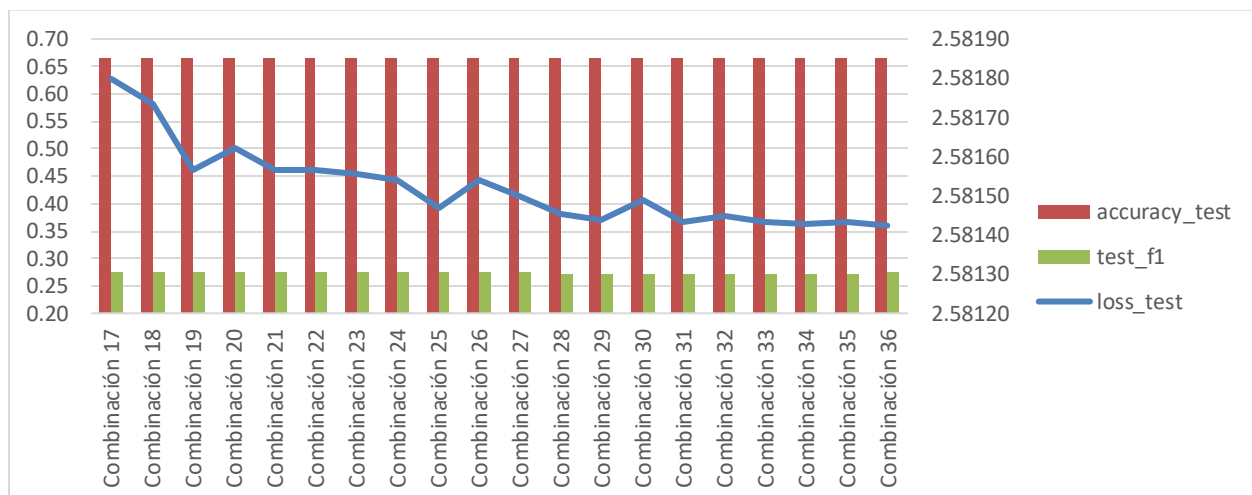
posibles combinaciones y se asignan variables a las métricas para llevar registro de los resultados a medida que corre el modelo.

Se toman las tasas de aprendizaje, incluyendo 2e-5, 3e-5, 5e-5 y 1e-4, combinadas con los demás hiper parámetros, y se analizan los resultados obtenidos (véase **Error! Reference source not found.**

Tabla 9. Hiper parámetros para los modelos considerados de Pysentimiento.

| 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 32. Resultados de los parámetros



Para el conjunto de entrenamiento, se observa que la pérdida más baja se obtiene en la combinación 17, que corresponde a los siguientes parámetros:

- lr (tasa de aprendizaje): 3e-5
- Épocas: 5

- Warmup ratio: 0.08

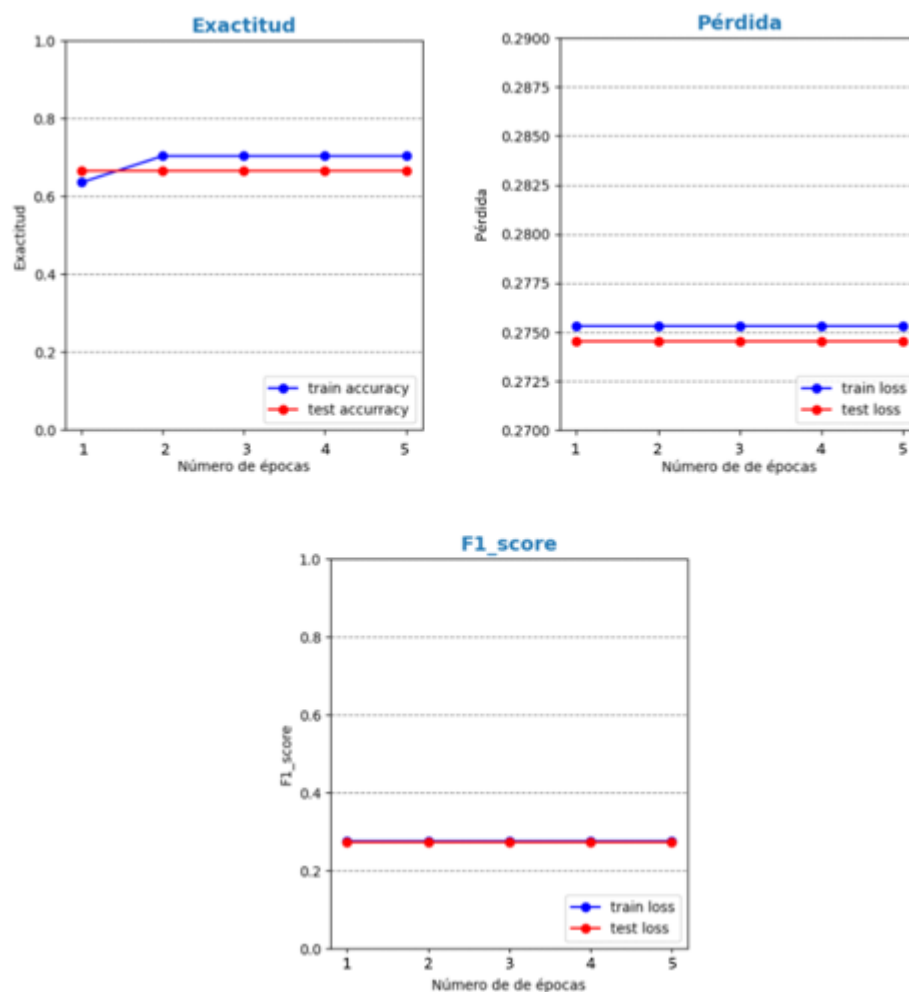
En cuanto a la exactitud, se encuentran valores similares para todas las combinaciones, excepto para aquellas con una tasa de aprendizaje de $1e-4$. Por otro lado, los valores del F1-score no varían en ninguna combinación.

Para el conjunto de prueba, se registra la pérdida más baja en la última combinación, utilizando los siguientes hiperparámetros: lr: $1e-4$, epochs: 5 y wr: 0.1. Sin embargo, tanto la medida de exactitud como la de F1-score son similares para todas las combinaciones.

Finalmente se probó el modelo con la combinación de mejor rendimiento, en este caso la número 17 (ver **Error! Reference source not found.**), teniendo en cuenta que su pérdida en el conjunto de entrenamiento fue la más baja. Se obtienen las gráficas de las métricas comparando el conjunto de entrenamiento y el de prueba.

Tabla 10. *Mejor combinación*

| Combinación de hiper parámetros | lr | batch_size | epochs | weight_decay | warmup_ratio | loss_train | accuracy_train | train_f1 | loss_test | accuracy_test | test_f1 |
|---------------------------------|-------|------------|--------|--------------|--------------|---------------------------|----------------|-----------------|-----------------|---------------|----------|
| Combinación 17 | 3E-05 | 32 | 5 | 0.10 | 0.08 | 7.58 932 | 0.7035 7 | 0.2 753 3 | 2.5 818 0 | 0.67 | 0.2 7 |

Figura 33. Métricas de los conjuntos de entrenamiento y prueba.

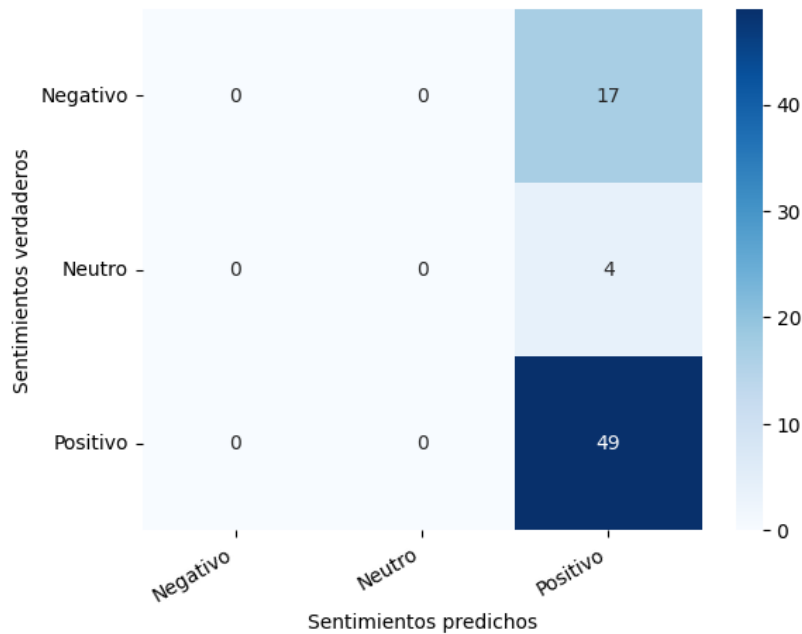
Se presenta un sub-ajuste en el modelo, que indica que el modelo no es generalizable al problema de análisis de sentimiento objeto de interés en esta investigación.

Una vez identificada la mejor combinación, se lleva a cabo el análisis de sentimiento para comparar el rendimiento de las predicciones del modelo con la etiquetación manual. En este caso, a diferencia de lo experimentado con las librerías, se realiza el análisis únicamente de los contenidos de los comentarios, ya que estos contienen más información para clasificar el

sentimiento. Como se ha mencionado, los títulos muchas veces estaban compuestos por una sola palabra.

Al observar la matriz de confusión (**Error! Reference source not found.**) 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 34. Matriz de confusión del etiquetado manual VS BERT para los contenidos



Acorde con lo anterior, se presentan las métricas de desempeño del modelo para este caso particular en la Tabla 10. El modelo clasificó todos los comentarios como positivos, lo que resulta

en un recall de 1.0. En otras palabras, identificó correctamente la totalidad de los comentarios positivos en comparación con el total clasificado en esa categoría (en este caso, todos los comentarios). Sin embargo, como se ha discutido previamente, esto no necesariamente indica que el modelo sea efectivo en la predicción de la polaridad positiva, sino más bien que es la única categoría que logra identificar. Por tanto, este modelo con la configuración y conjuntos de datos utilizado no tiene mejor desempeño que los modelos probados para la construcción de pysentimiento.

Tabla 11. Métricas del modelo BERT para el análisis de sentimiento

| Manual VS BERT | | | | | | |
|-----------------------|------------------|---------------|-----------------|----------------|-----------------|------------|
| | Precisión | Recall | F1-score | Support | Accuracy | MCC |
| Negativo | 0.00 | 0.00 | 0.00 | 17 | 0.7 | 0.00 |
| Neutro | 0.00 | 0.00 | 0.00 | 4 | | |
| Positivo | 0.70 | 1.00 | 0.82 | 49 | | |

7. Utilidad del análisis de sentimientos en empresas hoteleras

En este capítulo, se divide la base de datos en dos grupos de hoteles para hacer recomendaciones más específicas sobre los aspectos identificados, usando los contenidos de los comentarios. Dado que el Hotel Hampton by Hilton y el Dann Carlton, tienen 309 comentarios, lo que representa más del 80% de la base de datos.

Inicialmente, en la Figura 35. *Gráfica de barras de la cantidad de comentarios por categoría del primer grupo de hoteles* se evidencia que el hotel Dann aporta la mayor cantidad de comentarios y que el 72% de sus comentarios son positivos. De manera similar, para el hotel Hampton, el 71% de sus comentarios son positivos. Estos dos hoteles ocupan el tercer y quinto lugar en la plataforma de TripAdvisor en el área de Bucaramanga.

Figura 35. Gráfica de barras de la cantidad de comentarios por categoría del primer grupo de hoteles

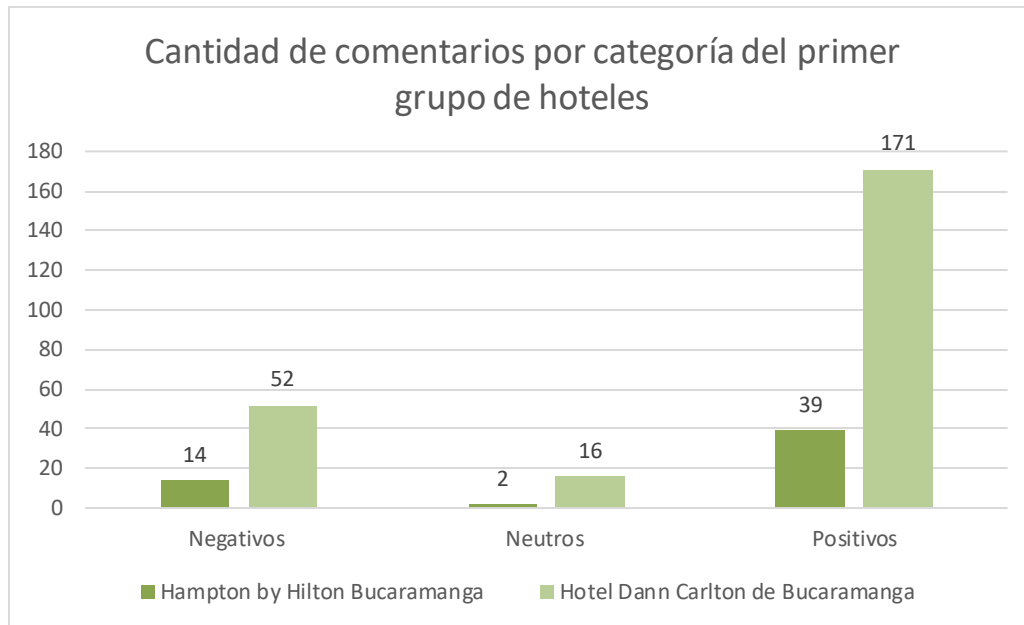
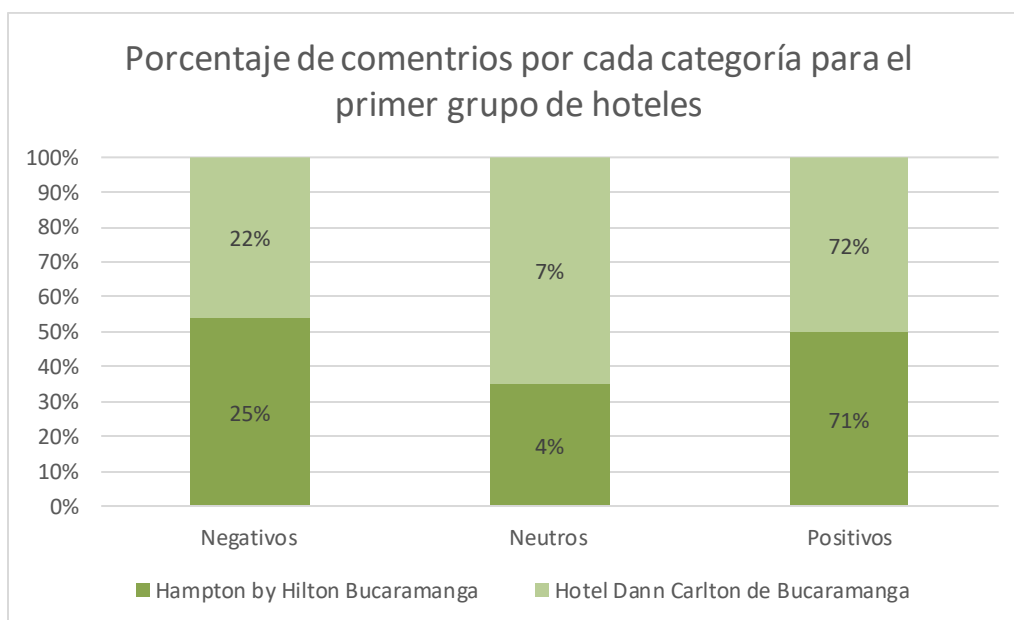


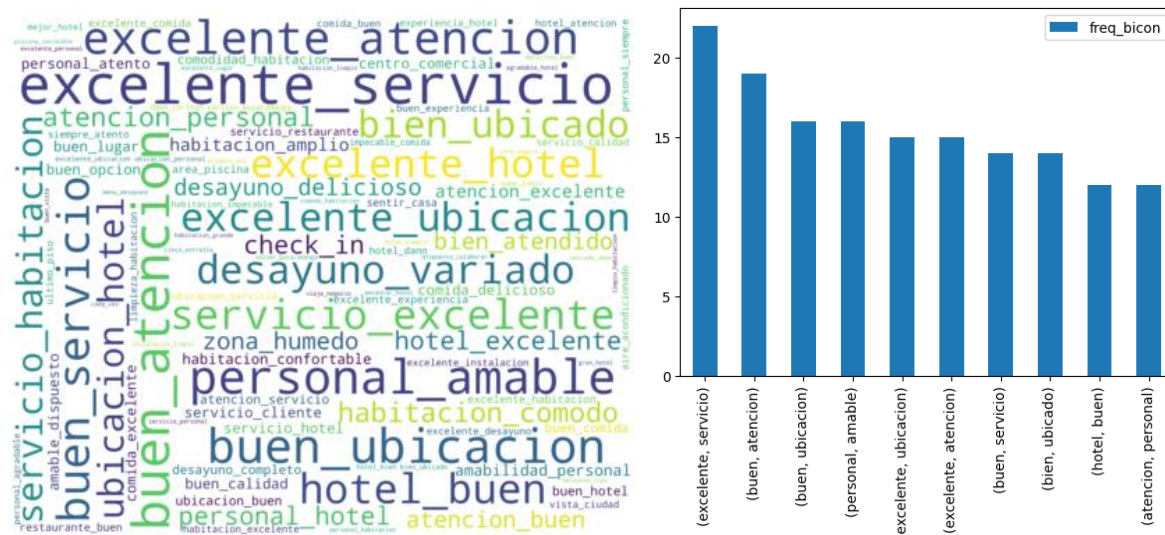
Figura 36. Gráfica de barras del porcentaje de comentarios por categoría del primer grupo de hoteles



amable por parte del personal del hotel puede motivar a los usuarios a dejar una reseña positiva tras su estadía.

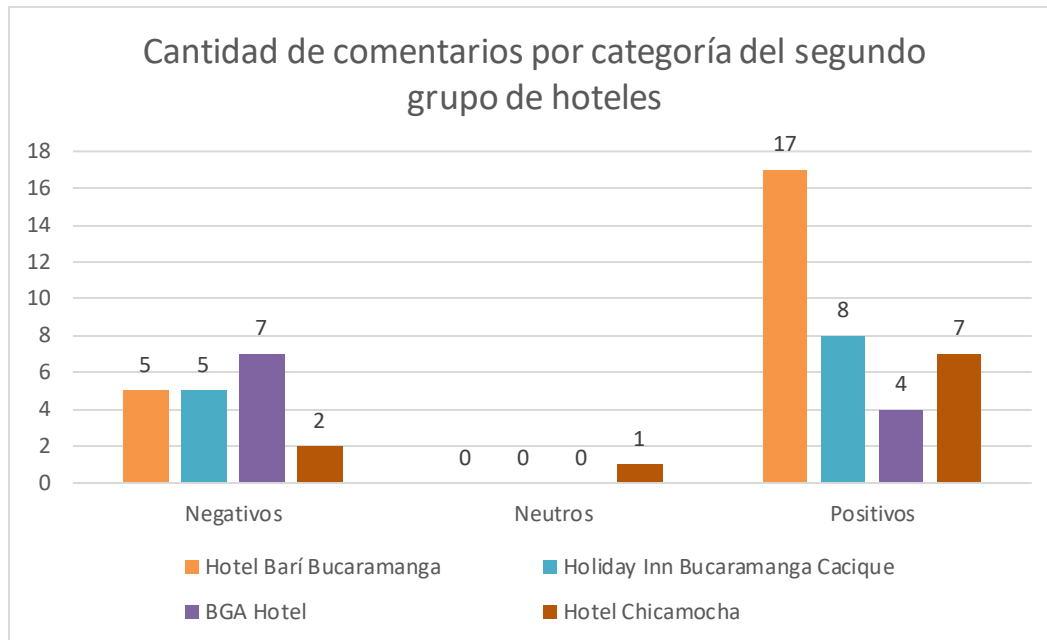
Igualmente, esto concuerda con lo que se evidenció durante la etiquetación manual. Por ejemplo, si en un comentario se mencionaba un inconveniente, se hacía énfasis en qué tanta atención le presta a dicho inconveniente, cómo reaccionan y si buscan resolverlo rápidamente. Cuando esto pasaba, los usuarios solían dejar una calificación positiva para el hotel.

Figura 39. Nube de palabras y diagrama de frecuencias de comentarios positivos para el primer grupo.



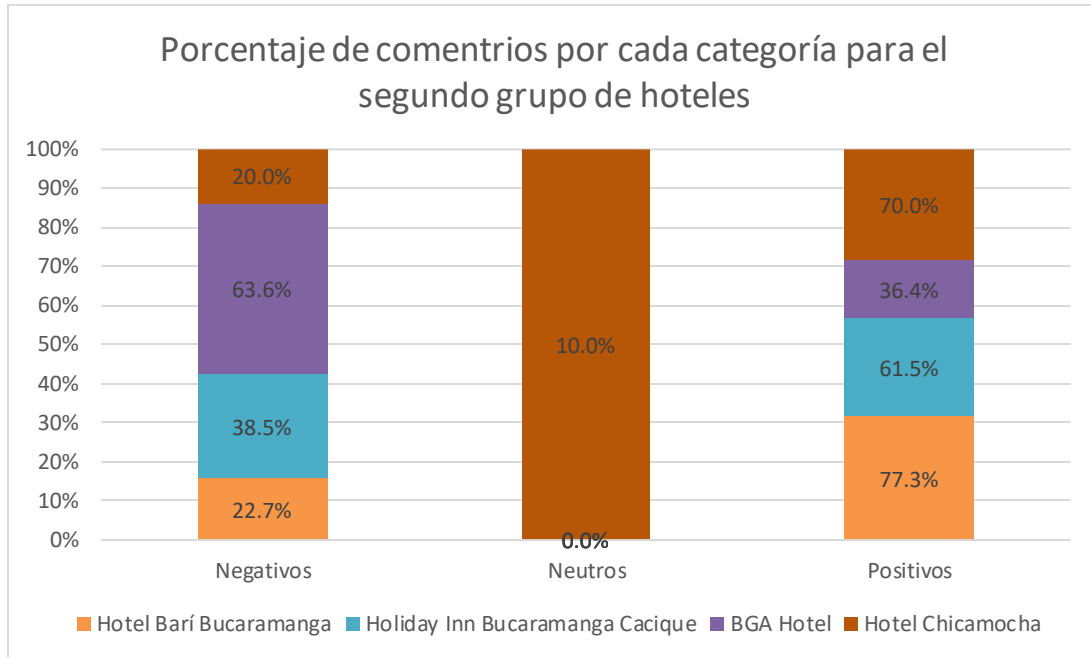
El segundo grupo se divide en cuatro hoteles, que representan el 12% de los datos con 56 comentarios, como se muestra en la Figura 40. De este grupo pequeño, resalta que solo se tiene un comentario neutro, correspondiente al Hotel Barí que tiene cerca del 40% de los comentarios en este caso. Sin embargo, no es el que mayor cantidad de comentarios tiene en cada categoría, ya que BGA Hotel tiene más comentarios negativos, a pesar de tener solo la mitad de los comentarios del Hotel Barí.

Figura 40. Gráfica de barras de la cantidad de comentarios por categoría del segundo grupo de hoteles



En concordancia con la idea anterior, BGA Hotel es el único establecimiento que tiene más comentarios negativos que positivos (Figura 44)

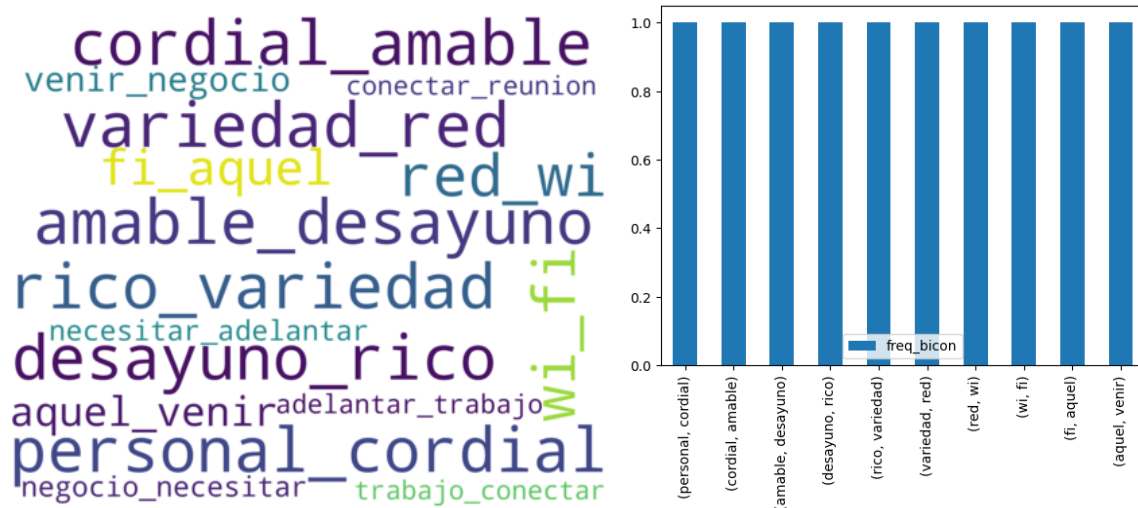
Figura 41. Gráfica de barras del porcentaje de comentarios por categoría del segundo grupo de hoteles



Comenzando con los comentarios negativos, a pesar de ser un grupo más pequeño, los bigramas más frecuentes dan más información. Al igual que en el primer grupo, se encuentra “aire acondicionado”, también “habitación aire” (ver Figura 42), por lo que, a pesar de que las frecuencias son bajas, si se suman a los comentarios del primer grupo, se puede concluir que varios usuarios tuvieron inconvenientes con el funcionamiento del aire. La palabra “habitación” aparece dos veces más acompañada de “después” y “mal”, la frecuencia es tan solo de 2 para estos términos, pero teniendo en cuenta que en este grupo solo hay 19 comentarios negativos, se puede decir que es importante tenerlo en cuenta. También se podrían interpretar los bigramas “señora recepción” y “pedir disculpas” como inconvenientes con el personal, que como se evidenció, es uno de los aspectos más importantes para el huésped.

Por último, se identifica el aspecto del restaurante y el aseo.

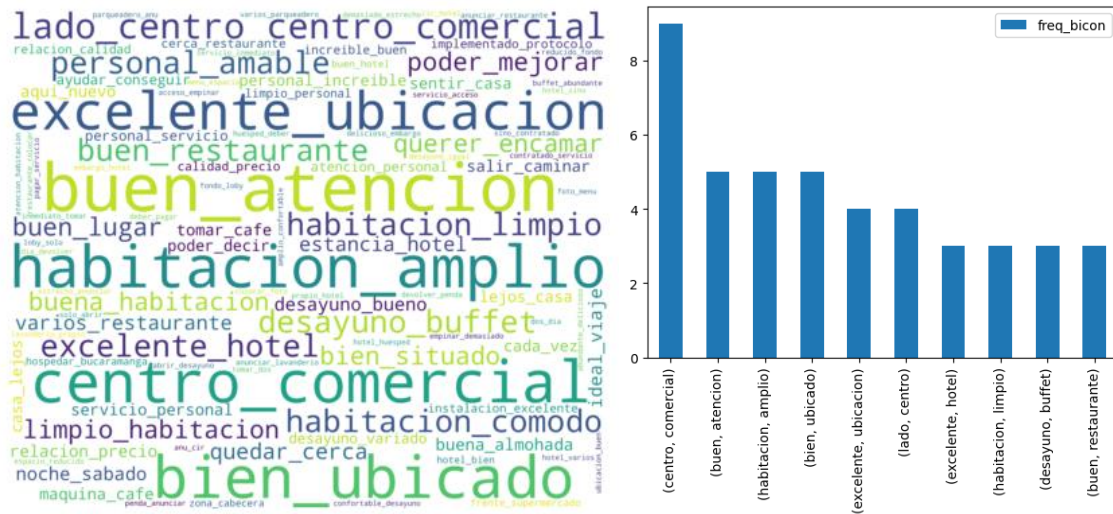
Figura 43. Nube de palabras y diagrama de frecuencias de comentarios neutros para el segundo grupo de hoteles



Para los comentarios positivos (ver Figura 44), el bigrama más frecuente es “centro comercial” y puede relacionarse con “lado centro” (indicando que está al lado de un centro comercial) “excelente ubicación” y “bien ubicado”. Hace referencia a las habitaciones en dos ocasiones, indicando que son amplias y limpias, también a la comida con “desayuno buffet” y “buen restaurante”.

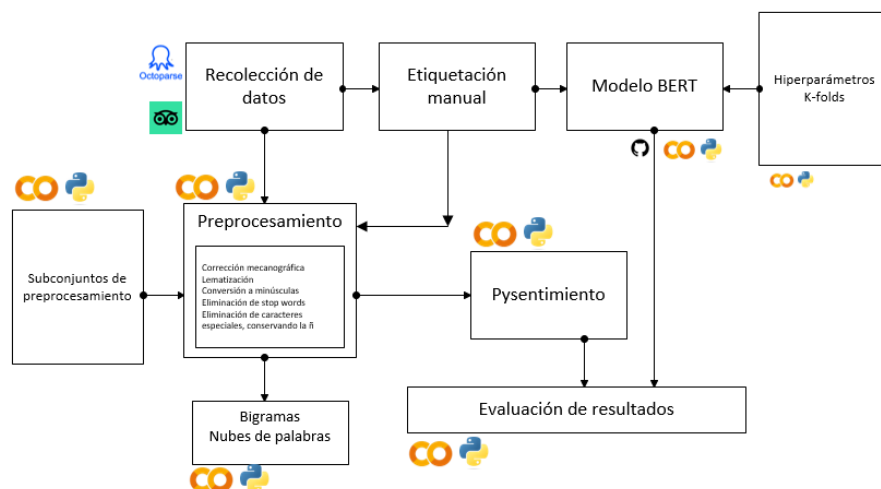
A diferencia del primer grupo, solo un bigrama hace referencia a la atención.

Figura 44. Nube de palabras y diagrama de frecuencias de comentarios positivos para el segundo grupo de hoteles.



Para finalizar, en la Figura 45 se presenta el esquema de la metodología usada en el proyecto, que puede ser escalada a una base de datos mucho más grande, reconociendo el análisis de sentimiento como una herramienta útil para la generación de información enriquecedora para los establecimientos hoteleros del sector.

Figura 45. Esquema de la metodología del proyecto



8. 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, 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 hiperparámetros, sin embargo, el rendimiento fue insatisfactorio debido a la limitada cantidad de comentarios y al desbalance de clases.

Finalmente, se realizan análisis adicionales al centrarse por separado en los dos grupos de hoteles. Este análisis revela que los comentarios negativos de los dos primeros hoteles se enfocan principalmente en el aire acondicionado y el proceso de registro en el hotel, mientras que los neutros mencionan la necesidad de aumentar la variedad, sin embargo, nada en concreto. En contraste, los comentarios positivos resaltan 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 proporcionan información más detallada, mencionando nuevamente el aire acondicionado y agregando el aspecto de las habitaciones y el servicio; mientras que los neutros indican problemas con la conexión wifi en uno de los hoteles. En los comentarios positivos, se valoran 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.

9. Limitaciones

La dificultad para encontrar librerías de análisis de sentimiento multiclase en español limita la cantidad de clasificadores con los que se realiza el experimento. En segundo lugar, la cantidad limitada de comentarios debido a la restricción de las fechas y en general al volumen limitado de datos en los hoteles TripAdvisor para la ciudad de Bucaramanga, fue un inconveniente al momento de entrenar el modelo, adicionalmente las clases desbalanceadas afectan el desempeño del modelo. Al mismo tiempo la capacidad del equipo de cómputo dificulta la ejecución del modelo, con largos periodos de espera y desconexiones inesperadas del entorno de ejecución.

10. Recomendaciones

Se recomienda en primer lugar, ampliar la base de datos bien sea ampliando el objeto de estudio más allá de la ciudad de Bucaramanga, o incluyendo otras plataformas como Booking, Yelp o Airbnb. Adicionalmente, aplicar métodos para tratar el problema del desbalance de clases en el conjunto de datos. Es posible explorar otras arquitecturas de modelo y plantear un problema de clasificación binario para enfocarse únicamente en los comentarios negativos y positivos. Respecto al análisis de sentimiento, sería ideal un grupo de expertos para la clasificación de los comentarios, o un grupo de personas entrenadas en el tema, y se podría explorar un enfoque de análisis de sentimientos basado en aspectos (ABSA).

Referencias Bibliográficas

- Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research*, 58(2), 175–191. <https://doi.org/10.1177/0047287517747753>
- Alosaimi, S., Alharthi, M., Alghamdi, K., Alsubait, T., & Alqurashi, T. (2020). Sentiment analysis of arabic reviews for Saudi hotels using unsupervised machine learning. *Journal of Computer Science*, 16(9), 1258–1267. <https://doi.org/10.3844/jcssp.2020.1258.1267>
- Ardila, C. (2022). *Prototipo de un Clasificador de Sentimientos para Chats de Atención al Cliente en Canales Digitales del Sector Salud*.
- Arroni, S., Galán, Y., Guzmán-Guzmán, X., Núñez-Valdez, E. R., & Gómez, A. (2023). Sentiment Analysis and Classification of Hotel Opinions in Twitter With the Transformer Architecture. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(1), 53–63. <https://doi.org/10.9781/ijimai.2023.02.005>
- Bang, T. S., & Sornlertlamvanich, V. (2018). Sentiment classification for hotel booking review based on sentence dependency structure and sub-opinion analysis. *IEICE Transactions on Information and Systems*, E101D(4), 909–916. <https://doi.org/10.1587/transinf.2016IIP0038>
- Benlahbib, A., & Nfaoui, E. H. (2020). Aggregating customer review attributes for online reputation generation. *IEEE Access*, 8, 96550–96564. <https://doi.org/10.1109/ACCESS.2020.2996805>
- Boshnjakoska, R. (2023, May). *Where is TripAdvisor Going? 39+ Signpost Statistics*.
- Buduma, N., & Locascio, N. (2017). *Fundamentals of Deep Learning*. O'Reilly.

- Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). A Practical Guide to Sentiment Analysis. En: Hussain, A. y Cambria, E. *Socio-Affective Computing*. https://doi.org/10.1007/978-3-319-55394-8_1
- Chang, Y. M., Chen, C. H., Lai, J. P., Lin, Y. L., & Pai, P. F. (2021). Forecasting hotel room occupancy using long short-term memory networks with sentiment analysis and scores of customer online reviews. *Applied Sciences (Switzerland)*, 11(21). <https://doi.org/10.3390/app112110291>
- Das, S., Mondal, S., Puri, V., & Vrana, V. (2022). Structural review of relics tourism by text mining and machine learning. *Journal of Tourism, Heritage and Services Marketing*, 8(2), 25–34. <https://doi.org/10.5281/zenodo.7358349>
- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://github.com/tensorflow/tensor2tensor>
- Dina, N. Z. (2020). Tourist sentiment analysis on TripAdvisor using text mining: A case study using hotels in Ubud, Bali. In *African Journal of Hospitality, Tourism and Leisure* (Vol. 9, Issue 2). <http://www.ajhtl.com>
- Equipo de administración Tripadvisor. (2019). *Acerca de Tripadvisor*. <https://tripadvisor.mediaroom.com/CO-about-us>
- Espinosa, E., Quiroga, M., & Rodriguez, C. (2023). *ProyectoMAD*. <https://github.com/moniiandrea/ProyectoMAD>
- Gárces Chaparro, T. I. (2019). *Análisis de sentimientos en redes sociales orientado a la percepción de la calidad de servicios de internet, redes móviles, tv cable y electricidad*. (Tesis de Pregrado). Universidad Andrés Bello, Santiago de Chile, Chile.

- Hameed, R. A., Abed, W. J., & Sadiq, A. T. (2023). Evaluation of Hotel Performance with Sentiment Analysis by Deep Learning Techniques. *International Journal of Interactive Mobile Technologies*, 17(9), 70–87. <https://doi.org/10.3991/ijim.v17i09.38755>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining* (Third Edition). Elsevier.
- Hernandez-Orallo, J. (2005). Knowledge Discovery from Databases. In *Encyclopedia of Database Technologies and Applications*. <https://dmip.webs.upv.es/papers/J.Hdez-Orallo-KDD-2.2-BW-033.pdf>
- Hutto, C. J., & Gilbert, E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. <http://sentic.net/>
- Intiaz, N., Ahmed, T., & Paul, A. (2021). Incentivized comment detection with sentiment analysis on hotel reviews. In *Asia-Pacific Journal of Science and Technology*. <https://www.tci-thaijo.org/index.php/APST/index>
- Instituto Municipal de Cultura y Turismo de Bucaramanga. (2022). *Boletín de Turismo de Bucaramanga*. <https://imct.gov.co/wp-content/uploads/2023/05/BOLETIN-DE-TURISMO-2022-CONSOLIDADO.pdf>
- Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, 41. <https://doi.org/10.1016/j.cosrev.2021.100413>
- Jayanto, R., Kusumaningrum, R., & Wibowo, A. (2022). Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory. *International Journal of Advances in Intelligent Informatics*, 8(3), 391–403. <https://doi.org/10.26555/ijain.v8i3.691>

- Kaushik, A., & Naithani, S. (2016). A Comprehensive Study of Text Mining Approach. En: Zeeshan, K., Nawaz, T., Adnan, H., Maqsood, M. y Amin, T. *IJCSNS International Journal of Computer Science and Network Security* (Vol. 16, Issue 2).
- Khamphakdee, N., & Seresangtakul, P. (2021). Sentiment analysis for Thai language in hotel domain using machine learning algorithms. *Acta Informatica Pragensia*, 10(2), 155–171. <https://doi.org/10.18267/j.aip.155>
- Lee, M., Jeong, M., & Lee, J. (2017). Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: A text mining approach. *International Journal of Contemporary Hospitality Management*, 29(2), 762–783. <https://doi.org/10.1108/IJCHM-10-2015-0626>
- Li, Q., Li, S., Zhang, S., Hu, J., & Hu, J. (2019). A review of text corpus-based tourism big data mining. In *Applied Sciences (Switzerland)* (Vol. 9, Issue 16). MDPI AG. <https://doi.org/10.3390/app9163300>
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining* (G. Hirst, Ed.). Morgan Claypool Publishers.
- Luo, J., Huang, S., & Wang, R. (2021). A fine-grained sentiment analysis of online guest reviews of economy hotels in China. *Journal of Hospitality Marketing and Management*, 30(1), 71–95. <https://doi.org/10.1080/19368623.2020.1772163>
- Luo, J. M., Vu, H. Q., Li, G., & Law, R. (2021). Understanding service attributes of robot hotels: A sentiment analysis of customer online reviews. *International Journal of Hospitality Management*, 98. <https://doi.org/10.1016/j.ijhm.2021.103032>
- Mejova, Y. (2009). *Sentiment Analysis: An Overview*. University of Iowa.

Méndez Pineda, J. D., & Talero Sarmiento, L. H. (2021). *Un modelo para la predicción del movimiento del precio de las acciones del mercado bursátil basado en un análisis de sentimiento y datos históricos de la BVC* (tesis de pregrado). Universidad Industrial de Santander, Bucaramanga, Colombia.

Ministerio de Comercio, I. y T. (2022). *Turismo en armonía con la vida*. <https://www.mincit.gov.co/participa/consulta-ciudadana/20-12-2022-plan-sectorial-de-turismo-vfpp.aspx>

Ministerio de Comercio, I. y T. (2023, July 5). *Colombia es el Tercer País En América En Superar Las Cifras Prepandemia En Turismo, Según La Organización Mundial Del Turismo*. <https://www.mincit.gov.co/prensa/noticias/turismo/colombia-supera-cifras-prepandemia-en-turismo-omt#:~:text=julio%20de%202023-,Colombia%20es%20el%20tercer%20pa%C3%ADs%20en%20Am%C3%A9rica%20en%20superar%20las,6%20millones%20de%20viajeros%20internacionales>.

Miranda, C. H., Guzmán, J., & Salcedo, D. (2016). Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles. *Procesamiento de Lenguaje Natural*, 25–32.

Mishra, D. N., & Panda, R. K. (2021). How delightful is indian wellness tourism? A netnographic study. *Advances in Hospitality and Tourism Research*, 9(1), 132–156. <https://doi.org/10.30519/ahtr.784232>

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. *Proceedings of the 2nd International Conference on Knowledge Capture, K-CAP 2003*, 70–77. <https://doi.org/10.1145/945645.945658>

- Nicolau, J. L., Xiang, Z., & Wang, D. (2023). Daily online review sentiment and hotel performance. *International Journal of Contemporary Hospitality Management*. <https://doi.org/10.1108/IJCHM-05-2022-0594>
- Nielsen, M. A. (2015). *Neural Networks and Deep Learning* (Determination Press). <http://neuralnetworksanddeeplearning.com/index.html>
- Oliveira, A. S., Renda, A. I., Correia, M. B., & Antonio, N. (2022a). Hotel customer segmentation and sentiment analysis through online reviews: An analysis of selected European markets. *Tourism and Management Studies*, 18(1), 29–40. <https://doi.org/10.18089/tms.2022.180103>
- Oliveira, A. S., Renda, A. I., Correia, M. B., & Antonio, N. (2022b). Hotel customer segmentation and sentiment analysis through online reviews: An analysis of selected European markets. *Tourism and Management Studies*, 18(1), 29–40. <https://doi.org/10.18089/tms.2022.180103>
- Ounacer, S., Mhamdi, D., Ardchir, S., Daif, A., & Azzouazi, M. (n.d.). Customer Sentiment Analysis in Hotel Reviews Through Natural Language Processing Techniques. En: *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 14, Issue 1). www.ijacsa.thesai.org
- Özen, İ. A., & Özgül Katlav, E. (2023). Aspect-based sentiment analysis on online customer reviews: a case study of technology-supported hotels. *Journal of Hospitality and Tourism Technology*, 14(2), 102–120. <https://doi.org/10.1108/JHTT-12-2020-0319>
- Pérez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D. A., Luque, F., Alemany, L. A., & Martínez, M. V. (2021). *pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks*. <http://arxiv.org/abs/2106.09462>

- Puh, K., & Bagić Babac, M. (2023). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights*, 6(3), 1188–1204. <https://doi.org/10.1108/JHTI-02-2022-0078>
- Roy, G. (2023). Travelers' online review on hotel performance – Analyzing facts with the Theory of Lodging and sentiment analysis. *International Journal of Hospitality Management*, 111. <https://doi.org/10.1016/j.ijhm.2023.103459>
- Sehgal, R. (2019). Sentimental Analysis On Hotel Reviews Using Classic Approaches And A Smaller Data Set. *International journal of scientific & technology research*, 8(09). www.ijstr.org
- Similar web. (2024). *Sitios web Viajes y turismo más visitados en Colombia*. <https://www.similarweb.com/es/website/booking.com/vs/tripadvisor.com/#competitors>
- Sontayasara, T., Jariyapongpaiboon, S., Promjun, A., Seelpipat, N., Saengtabtum, K., Tang, J., & Leelawat, N. (2021). Twitter sentiment analysis of bangkok tourism during covid-19 pandemic using support vector machine algorithm. *Journal of Disaster Research*, 16(1), 24–30. <https://doi.org/10.20965/jdr.2021.p0024>
- Suryadi, D., & Imran, J. A. (2022). A Topic Modeling and Sentiment Analysis Approach for Benchmarking of Hotels Based on Online Reviews. *Industrial Engineering and Management Systems*, 21(4), 646–657. <https://doi.org/10.7232/iems.2022.21.4.646>
- Valdivia, A., Hrabova, E., Chaturvedi, I., Luzón, M. V., Troiano, L., Cambria, E., & Herrera, F. (2019). Inconsistencies on TripAdvisor reviews: A unified index between users and Sentiment Analysis Methods. *Neurocomputing*, 353, 3–16. <https://doi.org/10.1016/j.neucom.2018.09.096>

- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*.
- Win, M. N., Ravana, S. D., & Shuib, L. (2022). Sentiment attribution analysis with hierarchical classification and automatic aspect categorization on online user reviews. *Malaysian Journal of Computer Science*, 35(2), 89–110. <https://doi.org/10.22452/mjcs.vol35no2.1>
- Yu, T. (2020). *Sentiment analysis in restaurants on social media reviews: the case of Giethoorn restaurants* (Tesis de maestría). Instituto Universitario de Lisboa, Lisboa, Portugal.
- Yu, W., Cui, F., & Hou, Z. (2023). The evolution of consumers' demand for hotels under the public health crisis: opinion mining from online reviews. *Current Issues in Tourism*, 26(12), 1974–1990. <https://doi.org/10.1080/13683500.2022.2073204>
- Zaida De León Martín, Andrea García Hernández, & Esperanza Gil Soto. (2020). *Analysis of hotel response management to online reviews* (Tesis de pregrado). Universidad de La Laguna, San Cristóbal de La Laguna, España.
- Zapata, A. (2023, June 18). *Dominando la Matriz de Confusión: La guía completa para entender el rendimiento de nuestros modelos de Machine Learning*. <https://es.linkedin.com/pulse/dominando-la-matriz-de-confusi%C3%B3n-gu%C3%ADa-completa-para-el-zapata-guzman>
- Zhang, J., Lu, X., & Liu, D. (2021). Deriving customer preferences for hotels based on aspect-level sentiment analysis of online reviews. *Electronic Commerce Research and Applications*, 49. <https://doi.org/10.1016/j.elerap.2021.101094>