ULPGC Universidad de Las Palmas de Gran Canaria | Escuela de Ingeniería Informática

DOBLE GRADO EN INGENIERÍA INFORMÁTICA Y ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Tratamiento y análisis de las opiniones de los clientes de hoteles con técnicas de Inteligencia Artificial y extracción de información relevante para la gestión

Presentado por: Ana Carla Morales Padrón

DNI: 45348999-Z

Tutorizado por:

Francisco Mario Hernández Tejera

Jacques Bulchand Gidumal

En Las Palmas de Gran Canaria, a 23 de julio de 2021

AGRADECIMIENTOS

A mis tutores académicos, Mario y Jacques, por confiar y darme la oportunidad de plasmar esta idea, así como guiarme a lo largo de la elaboración de este trabajo. Además, me gustaría hacer especial mención a José Daniel, quien se ha preocupado por los alumnos de esta titulación desde el minuto uno, ayudándonos y proporcionando soluciones a los diferentes obstáculos que nos hemos ido encontrando a lo largo de estos años. Gracias.

A mi gran familia, por estar pendientes de cada paso que daba y confiar en mí desde el inicio de esta aventura.

A mis padres, porque sin ellos no sería la persona que soy hoy en día. Por enseñarme que no estoy sola en esto y que siempre estarán ahí para ayudar a levantarme. Gracias por ser un apoyo incondicional en esta etapa de mi vida y enseñarme a no rendirme.

A Lara, mi persona favorita. Por transmitirme su alegría y positividad, y creer en mí más que yo misma.

"You will never walk alone"

RESUMEN

La segunda generación de servicios basados en la web se caracteriza por tener un contenido generado por el consumidor (CGC), que permite a los usuarios compartir, crear o intercambiar información e ideas en comunidades y redes virtuales. Hoy en día, los medios sociales desempeñan un papel muy importante en casi todas las áreas, especialmente en la industria del turismo. Es por ello, que este trabajo analiza el CGC en TripAdvisor, con un estudio de caso sobre el turismo de la isla de Gran Canaria. Se analizan las reseñas de los diferentes alojamientos turísticos con el fin de explorar cómo los usuarios perciben sus servicios e instalaciones. Esta información ayudará a los proveedores de servicios turísticos a centrarse en mejores prácticas, incluyendo cambios en el modelo de negocio, nuevas estrategias de marketing, mejora de servicios o infraestructuras, e incluso, un análisis de la competencia.

Palabras clave: NLP, análisis de sentimientos, turismo, reseñas, *web scraping*, ABSA

ABSTRACT

The second generation of web-based services is characterised by user-generated content (UGC), which allows users to share, create or exchange information and ideas in virtual communities and networks. Nowadays, social media plays a very important role in almost all fields, especially in the tourism industry. For this reason, this work examines UGC on TripAdvisor, with a case study on the island of Gran Canaria. The reviews of different tourist accommodations are analysed in order to explore how users perceive their services and facilities. This information will help tourism service providers to focus on best practices, including changes in the business model, new marketing strategies, improvement of services or infrastructures, and even an analysis of the competition.

Keywords: NLP, sentiment analysis, tourism, reviews, web scraping, ABSA

GUÍA DE LECTURA

A continuación, se especifican aquellos capítulos y epígrafes recomendados para la lectura de los distintos tribunales. No obstante, se sugiere la revisión del conjunto de este trabajo para obtener una visión íntegra del trabajo realizado.

Tribunal del Grado en Administración y Dirección de Empresas

- Capítulo 1 – Introducción
- Capítulo 2 – Marco teórico (en especial, el epígrafe 2.1 *Los medios sociales*, el 2.2 *Las valoraciones de los productos y servicios en Internet*, y el 2.4 *CRISP-DM (Cross-Industry Standard Process for Data Mining)*)
- Capítulo 3 – Objetivo
- Capítulo 4 – Metodología
- Capítulo 7 – Resultados
- Capítulo 8 – Uso de la herramienta desarrollada en el entorno empresarial
- Capítulo 9 – Conclusiones

Tribunal del Grado en Ingeniería Informática

- Capítulo 1 – Introducción
- Capítulo 2 – Marco teórico (en especial, el epígrafe 2.2 *Métodos y técnicas de análisis*, y el 2.4 *CRISP-DM (Cross-Industry Standard Process for Data Mining)*)
- Capítulo 3 – Objetivo
- Capítulo 4 – Metodología
- Capítulo 5 – Recursos Tecnológicos
- Capítulo 6 – Desarrollo
- Capítulo 7 – Resultados
- Capítulo 9 – Conclusiones

ÍNDICE DE CONTENIDOS

ÍNDICE DE TABLAS

ÍNDICE DE ILUSTRACIONES

ÍNDICE DE ECUACIONES

1. INTRODUCCIÓN

Actualmente vivimos en una época en la que la información disponible es cada vez mayor. El volumen de datos aumenta considerablemente año a año, haciendo que el procesamiento de estos sea cada vez más complicado.

El rápido avance en las tecnologías y sistemas de información ha contribuido a que las empresas puedan explotar y dar un buen uso a esos datos, pudiendo extraer información valiosa y útil para la toma de decisiones y mejora de los servicios.

Este trabajo busca, mediante la ejecución de técnicas y algoritmos de inteligencia artificial, analizar la información relevante y necesaria para mejorar la gestión empresarial de los alojamientos turísticos en Gran Canaria.

El estudio comprende la extracción de valoraciones hechas por los turistas durante el año 2019 en la plataforma TripAdvisor con el fin de analizar la opinión que tienen acerca de un alojamiento turístico, los aspectos que más y menos valoran, así como las prestaciones, instalaciones y servicios que ofertan. Se determinó este año de estudio debido al impacto que generó la COVID-19 en los datos del año 2020.

La extracción de estas opiniones se llevó a cabo mediante la técnica *web scraping*, para su posterior tratamiento con algoritmos de traducción y análisis de sentimientos (*sentiment analysis*). Una vez obtenidos los resultados, elaboraremos un cuadro de mandos para visualizar la información recabada y resultante del proyecto. No obstante, serán los propios hoteles quienes podrán hacer uso de esta herramienta y utilizar el informe generado para realizar una mejora de su gestión e, incluso, efectuar un análisis de la competencia.

En este epígrafe, se presentarán los objetivos y motivación inicial para la realización de este trabajo. Seguidamente, se incluyen las competencias académicas cubiertas durante el desarrollo del proyecto; así como la estructura que sigue este estudio. Por último, se presenta una comparación entre la temporalización establecida inicialmente, y la que se ha seguido realmente, debido a las incidencias ocurridas durante la elaboración del trabajo.

1.1 Objetivos y motivación del trabajo

El objetivo del trabajo consiste en analizar las valoraciones de los turistas con el fin de recolectar información relevante y útil para tomar decisiones de mejora en la gestión empresarial y hotelera.

Las tareas llevadas a cabo en el proyecto se centran en automatizar el proceso de recolección de reseñas de los alojamientos turísticos de Gran Canaria durante el año 2019, traducirlas a un idioma común, y realizar un análisis de sentimientos a nivel de aspecto. Esto es una técnica de análisis de textos que clasifica los datos por aspectos e identifica el sentimiento atribuido a cada uno de ellos. Desde el punto de vista empresarial, las compañías pueden utilizarlo para comprender a sus clientes a un nivel más profundo, es decir, obtener información relevante y específica de ellos, de forma totalmente automatizada. En este trabajo en específico, se pretende descubrir aquellas características de los alojamientos que los turistas valoran de forma positiva y negativa para que, con la información obtenida, los gestores hoteleros traten de mitigar sus debilidades y realzar sus fortalezas. Esto se puede traducir en cambios en sus estrategias o modelos de negocio, con el objetivo de mejorar su actividad empresarial y ofrecer un servicio que garantice la satisfacción de sus clientes.

Es decir, se podrán valorar opiniones de forma masiva en cualquier idioma para que los hoteles puedan analizar el estado de sus servicios, instalaciones y gestión. De esta forma, se podrá obtener una visión general del alojamiento, pudiendo comparar los resultados con los de la competencia o realizar un análisis por zonas geográficas.

1.2 Estructura del trabajo

La memoria se divide en siete epígrafes. El presente se basa en obtener una visión general de la temática del trabajo, la temporalización y planificación, así como los objetivos del mismo.

El siguiente capítulo está dedicado a la investigación y estudio previo relacionado con los retos y temas que se abordan en el trabajo. En él se incluyen la importancia que tiene la recopilación de datos y valoraciones dentro del contexto turístico, así como el uso de los medios sociales para las empresas y su impacto.

El tercer epígrafe está dedicado a la metodología seguida en la realización del proyecto. Se incluyen las diferentes etapas por las que ha pasado el proyecto, como son la planificación del trabajo, la descarga de datos, su formateo, transformaciones y posterior análisis.

Seguidamente, se abordan los recursos tecnológicos utilizados en la elaboración del proyecto, tanto software como hardware, con su debida justificación en su elección. Además, se incluyen los algoritmos y paquetes utilizados en el desarrollo del trabajo.

El capítulo quinto comprende el desarrollo e implementación de todo el proceso donde se explotaron los modelos y métodos mencionados en el apartado tres.

Tras este apartado, se detallan los resultados obtenidos en el capítulo seis.

En el siguiente capítulo se describe el uso que se le puede dar al software creado con la explicación de varios casos de uso a modo de ejemplo.

A continuación, encontramos las conclusiones a las que se llegan tras la realización del proyecto, así como las principales aportaciones y los trabajos futuros.

Posteriormente, en el capítulo nueve, se citan aquellas referencias bibliográficas utilizadas durante el desarrollo del trabajo, ya sean correspondientes al marco teórico, como información de las herramientas/algoritmos utilizados.

En la parte final, encontramos los Anexos que recogen el correspondiente enlace a la visualización del proceso del presente trabajo.

Por último, cabe mencionar que, en la redacción del documento, y en lo posible, se ha planteado entre paréntesis la traducción de ciertos términos en inglés. Sin embargo, hay casos en los que se traduce la primera vez, pero se sigue usando en inglés por la extensión de su uso en el ámbito del trabajo o porque son acrónimos muy extendidos, como por ejemplo es el caso de CRISP-DM.

1.3 Justificación de las competencias cubiertas

En este apartado se recogen aquellas competencias académicas que han estado presentes y en práctica durante el desarrollo del trabajo, así como la justificación de las mismas.

Se comenzará citando las competencias generales de la Universidad de Las Palmas de Gran Canaria para, posteriormente, exponer las competencias propias de la titulación del Grado de Ingeniería Informática y del Grado en Administración y Dirección de Empresas.

1.3.1 Competencias nucleares Universidad de Las Palmas de Gran Canaria

N1.- Comunicarse de forma adecuada y respetuosa con diferentes audiencias (clientes, colaboradores, promotores, agentes sociales, etc.), utilizando los soportes y vías de comunicación más apropiados (especialmente las nuevas tecnologías de la información y la comunicación) de modo que pueda llegar a comprender los intereses, necesidades y preocupaciones de las personas y organizaciones, así como expresar claramente el sentido de la misión que tiene encomendada y la forma en que puede contribuir, con sus competencias y conocimientos profesionales, a la satisfacción de esos intereses, necesidades y preocupaciones.

Durante el desarrollo del trabajo fue necesaria y esencial la comunicación con ambos tutores.

N2.- Cooperar con otras personas y organizaciones en la realización eficaz de funciones y tareas propias de su perfil profesional, desarrollando una actitud reflexiva sobre sus propias competencias y conocimientos profesionales y una actitud comprensiva y empática hacia las competencias y conocimientos de otros profesionales.

Esta competencia queda cubierta por la cooperación y trabajo continuo con los tutores académicos.

N3.- Contribuir a la mejora continua de su profesión, así como de las organizaciones en las que desarrolla sus prácticas a través de la participación activa en procesos de investigación, desarrollo e innovación.

La realización de este trabajo lleva la obtención y justificación de esta competencia.

1.3.2 Competencias del Grado de Ingeniería Informática

En primer lugar, se identifican las competencias generales del Grado de Ingeniería Informática que han sido cubiertas:

T1.- Capacidad para concebir, redactar, organizar, planificar, desarrollar y firmar proyectos en el ámbito de la ingeniería en informática que tengan por objeto, de acuerdo con los conocimientos adquiridos según lo establecido en apartado 5 de la resolución indicada, la concepción, el desarrollo o la explotación de sistemas, servicios y aplicaciones informáticas. (G1, G2).

La redacción y elaboración de esta memoria justifica la competencia descrita.

T2.- Capacidad para dirigir las actividades objeto de los proyectos del ámbito de la informática, de acuerdo con los conocimientos adquiridos según lo establecido en apartado 5 de la resolución indicada. (G1, G2).

La realización de este trabajo conlleva la dirección y organización de un conjunto de actividades, por lo que su elaboración explica la consecución de la competencia anterior.

T4.- Capacidad para definir, evaluar y seleccionar plataformas hardware y software para el desarrollo y la ejecución de sistemas, servicios y aplicaciones informáticas, de acuerdo con los conocimientos adquiridos según lo establecido en apartado 5 de la resolución indicada. (G1, G2).

Para la correcta evaluación y análisis de los datos se llevó a cabo una selección de aquellas plataformas y aplicaciones más adecuadas según el objetivo que se quería conseguir.

T5.- Capacidad para concebir, desarrollar y mantener sistemas, servicios y aplicaciones informáticas empleando los métodos de la ingeniería del software como instrumento para el aseguramiento de su calidad, de acuerdo con los conocimientos adquiridos según lo establecido en apartado 5 de la resolución indicada. (G1, G2).

Una vez seleccionadas las plataformas software más adecuadas, se desarrollaron aplicaciones informáticas para poder realizar el análisis requerido por la investigación, por lo que esta competencia queda cubierta.

T8.- Conocimiento de las materias básicas y tecnologías, que capaciten para el aprendizaje y desarrollo de nuevos métodos y tecnologías, así como las que les doten de una gran versatilidad para adaptarse a nuevas situaciones. (G3, N3).

La mayoría de las técnicas y tecnologías aplicadas en el desarrollo del trabajo fueron implementadas por primera vez, por lo que requirieron de un aprendizaje y formación previa para poder ejecutarlas de forma correcta.

T9.- Capacidad para resolver problemas con iniciativa, toma de decisiones, autonomía y creatividad. Capacidad para saber comunicar y transmitir los conocimientos, habilidades y destrezas de la profesión de Ingeniero Técnico en Informática. (G4, N1).

La realización de este trabajo justifica la consecución de esta competencia.

A continuación, se exponen las competencias específicas:

CII01.- Capacidad para diseñar, desarrollar, seleccionar y evaluar aplicaciones y sistemas informáticos, asegurando su fiabilidad, seguridad y calidad, conforme a principios éticos y a la legislación y normativa vigente.

Esta competencia quedaría cubierta por los mismos motivos que las competencias generales T4 y T5, ya que en su desarrollo se cumple la normativa y legislación vigente.

CII06.- Conocimiento y aplicación de los procedimientos algorítmicos básicos de las tecnologías informáticas para diseñar soluciones a problemas, analizando la idoneidad y complejidad de los algoritmos propuestos.

Esta competencia se encuentra satisfecha con la elección de los algoritmos más adecuados y óptimos para el análisis de los datos.

CII07.- Conocimiento, diseño y utilización de forma eficiente los tipos y estructuras de datos más adecuados a la resolución de un problema.

Debido a la gran cantidad de datos extraídos para la elaboración del trabajo, se tuvo que utilizar estructuras de datos que se ajustaran de mejor manera y facilitaran el desarrollo de la metodología.

CII08.- Capacidad para analizar, diseñar, construir y mantener aplicaciones de forma robusta, segura y eficiente, eligiendo el paradigma y los lenguajes de programación más adecuados.

Esta competencia se encuentra cubierta teniendo en cuenta la selección de lenguajes de programación utilizados para el desarrollo del proyecto.

CII13.- Conocimiento y aplicación de las herramientas necesarias para el almacenamiento, procesamiento y acceso a los Sistemas de información, incluidos los basados en web.

El acceso y la conexión a las diferentes APIs para el análisis y la posterior elaboración de conclusiones, así como la descarga de datos iniciales de la plataforma de viajes TripAdvisor, justifican la consecución de esta competencia.

CII15.- Conocimiento y aplicación de los principios fundamentales y técnicas básicas de los sistemas inteligentes y su aplicación práctica.

La utilización de herramientas de Inteligencia Artificial para abordar los datos y sacar conclusiones justifica la consecución de esta competencia.

1.3.3 Competencias del Grado en Administración y Dirección de Empresas

Del mismo modo que en el apartado anterior, se expondrán en primer lugar las competencias generales del Grado en Administración y Dirección de Empresas:

CG1.- Capacidad de análisis y síntesis.

Esta competencia queda cubierta por el previo análisis, investigación y recapitulación de la bibliografía relacionada con la temática del trabajo, así como las metodologías utilizadas.

CG2.- Capacidad de organización y planificación.

La propia realización y planificación temporal de este trabajo justifica la consecución de esta competencia.

CG3.- Comunicación oral y escrita en lengua española.

Esta competencia se encuentra satisfecha con la elaboración de esta memoria, además de la presentación y defensa del trabajo descrito.

CG6.- Capacidad para la resolución de problemas.

La realización de este trabajo demuestra la consecución de esta competencia, pues durante su desarrollo, se abordaron diferentes problemas que tuvieron que ser solventados.

CG7.- Capacidad de tomar decisiones.

Durante la elaboración de este proyecto se llevaron a cabo múltiples decisiones, ya sea por la aparición de algún problema o la elección de metodologías/herramientas, cubriendo la competencia descrita anteriormente.

CG8.- Habilidades en la búsqueda, identificación, análisis e interpretación de fuentes de información diversas.

Esta competencia queda satisfecha con la búsqueda y tratamiento de las diferentes fuentes bibliográficas que se presentan en este trabajo.

CG24.- Defender un punto de vista mostrando y apreciando las bases de otros puntos de vista discrepantes.

La realización de este trabajo conlleva la consecución de esta competencia.

CG25.- Capacidad de aprendizaje autónomo.

Muchas de las herramientas utilizadas en la elaboración de este proyecto tuvieron que ser aprendidas anteriormente, al no poseer conocimientos suficientes de manera previa, justificando así la competencia anterior.

Para finalizar este epígrafe, se presentan a continuación, las competencias específicas:

CE1.- Capacidad de aplicar los conocimientos en la práctica.

La realización de este trabajo conlleva al logro de esta competencia.

CE2.- Habilidad para el diseño y gestión de proyectos.

La realización de este proyecto justifica la consecución de esta competencia, debido a la planificación y desarrollo del mismo.

CE3.- Habilidad de transmisión de conocimientos.

La presentación de las conclusiones y resultados obtenidos en el proyecto, justifican la competencia descrita.

1.4 Temporalización

La temporalización inicial del trabajo fue descrita y establecida inicialmente en el documento TFT01. En él se definieron las diferentes tareas que se iban a llevar a cabo en el proceso de elaboración del trabajo, con su respectiva estimación de horas. En un

principio se estipularon 300 horas totales, 150 horas para cada titulación. En la tabla 1 se presentan las tareas de manera más detallada, junto con la duración prevista para cada una.

En este tipo de trabajos existe un componente exploratorio que a veces produce, como en este caso, que la planificación original no se cumpla estrictamente y se produzcan incrementos en alguna o algunas partes, ya que van surgiendo diferentes problemáticas que hay que ir abordando.

En este caso, en la descarga de datos mediante *web scraping* y la implementación de traducción automática y análisis de sentimientos mediante IA *(Tarea 2.1 y Tarea 2.2)*, se emplearon más horas de las inicialmente estimadas, ya que se manejaron grandes cantidades de datos, con un volumen era superior al inicialmente previsto. Además, se empleó bastante tiempo en seleccionar y manejarse con la librería que se adecuara al objetivo del proyecto. En conjunto, estas actividades supusieron unas treinta horas más de las inicialmente consideradas.

Tabla 1. Temporalización inicialmente prevista de las tareas

Plan de trabajo: Se desglosará de manera detallada el trabajo del TFT en fases, con su duración estimada en horas (total 300). Cada fase, a su vez, se desglosará en tareas concretas; en el caso del doble grado deberá indicarse para cada tarea si es común o específica de alguno de los grados.

Fases	***Duración Estimada*** *(horas)*	***Tareas*** *(nombre y descripción, obligatorio al menos una por fase)*
Estudio previo / Análisis	40 (GADE) 10 (GII)	Tarea 1.1: Análisis de la importancia de las redes sociales para las empresas e impacto en sus resultados. Específica: GADE
		Tarea 1.2: Análisis del uso de las redes sociales en el contexto turístico. Específica: GADE
		Tarea 1.3: Análisis y elección de las herramientas informáticas adecuadas. Específica: GII
Diseño / Desarrollo / Implementación	60 (GADE) 90 (GII)	Tarea 2.1: Descarga de datos mediante web scraping. Específica: GII
		Tarea 2.2: Implementación de traducción automática y análisis de sentimientos mediante IA. Específica: GII
		Tarea 2.3: Tratamiento de los datos con los análisis seleccionados. Común
Evaluación / Validación / Prueba	20 (GADE) 30 (GII)	Tarea 3.1: Verificación del funcionamiento de la herramienta desarrollada. Común
		Tarea 3.2: Análisis de posibles mejoras o correcciones. Común
		Tarea 3.3: Propuesta de acciones de mejora en la gestión a partir de los resultados obtenidos. Específica: GADE
Documentación / Presentación	30 (GADE) 20 (GII)	Tarea 4.1: Redacción de la memoria. Común
		Tarea 4.2: Preparación de la defensa. Común

Fuente: Elaboración propia

2. MARCO TEÓRICO

Este epígrafe se comentará, por un lado, qué son los medios sociales y el impacto que generan en las empresas. Además, se describirá la importancia que tienen las valoraciones de productos y servicios en Internet y cómo afectan a la industria del turismo, con el caso particular de los alojamientos turísticos en TripAdvisor. Esta parte se corresponderá con el Grado de Administración y Dirección de Empresas.

Y, por el otro lado, hablaremos teóricamente del procesamiento de lenguaje natural y su metodología; y del análisis de sentimientos, junto con sus aplicaciones y limitaciones, con el caso particular del nivel más complicado y profundo que encontramos: el nivel de aspecto. Asimismo, se comentarán los métodos y modelos de las técnicas utilizadas en la comprensión del lenguaje, como son las redes neuronales o el modelo arquitectónico *Transformer*. Todo ello, referido al Grado de Ingeniería Informática.

Por último, definiremos, desde el punto de vista teórico, la metodología empleada en el desarrollo del presente trabajo, CRISP-DM, junto con las distintas tareas que comprende.

2.1 Los medios sociales

Con la llegada de la Web 2.0 han aparecido nuevos retos y oportunidades para alcanzar a más clientes o consumidores potenciales, gracias a la interactividad y formación de comunidades ofrecidos. Esta segunda generación de servicios se caracteriza por tener contenido generado por el consumidor (CGC), provocando que los usuarios puedan acceder y compartir información de una manera interactiva e informal. Es decir, se está produciendo una evolución en el modelo tradicional de marketing empresa-consumidor, hacia uno de intercambio de información entre iguales.

Esto es lo que se conoce como los medios sociales, plataformas online que permiten no sólo el acceso a cierto contenido sino la oportunidad para que los usuarios puedan colaborar e intercambiar información entre ellos, permitiendo que millones de personas puedan conectarse desde cualquier parte del mundo, y puedan acceder a la información disponible desde cualquier dispositivo (Bravo, 2018). Como mencionamos anteriormente, han supuesto un cambio en la forma de comunicación tradicional, permitiendo la creación de comunidades en línea con el fin de compartir información, intereses y otros contenidos (Yogesh y Yesha, 2014).

Algunos ejemplos de estos medios son los blogs, los foros, las redes profesionales o de empresa, los micro blogs, los juegos sociales, compartir vídeos o fotos, las redes sociales, los marcadores sociales, mundos virtuales, o crítica de productos o servicios (Aichner y Jacob, 2015), en la que profundizaremos más adelante. Como podemos ver, son actividades que implican el uso de la tecnología y la interacción social.

Una posible clasificación se presenta en la tabla 2, donde se diferencian los medios sociales atendiendo a cuatro categorías: de publicación, de colaboración, multimedia y de entretenimiento. Hemos incluido ejemplos de los medios sociales más conocidos entre paréntesis.

Tabla 2. Clasificación de medios sociales

Clasificación	
Medios sociales de publicación	- Redes sociales o comunidades virtuales (Facebook, LinkedIn, Instagram…) - Blogs (Blogger, WordPress, LiveJournal…) - Microblogging (Twitter, Tumblr…)
Medios sociales de colaboración	- Wiki (Wikipedia) - Marcadores sociales (Digg, Pinterest…) - Sitios de opinión (TripAdvisor, Yelp…)
Medios sociales multimedia	- Fotos (Flickr, Snapchat…) - Vídeos (YouTube, Twitch, TikTok…) - Música (SoundCloud…)
Medios sociales de entretenimiento	- Juegos en línea (World of Warcraft, The Sims Online…)

Fuente: Elaboración propia

El primero de ellos son los medios sociales para publicar, que comprenden aquellas plataformas en las que cualquier usuario, de cualquier parte del mundo puede hacer público y transmitir el contenido que quiera. A su vez, dentro de estos, encontramos las redes

sociales, comunidades formadas por diferentes usuarios y organizaciones que se conectan entre sí; los blogs y el microblogging, sitios para fomentar el descubrimiento, intercambio y comentarios acerca de un cierto contenido.

Por otro lado, encontramos los medios de colaboración, en los que existe una interacción y participación entre usuarios con el fin de obtener un objetivo común. Estos engloban los Wikis, espacios donde varios autores pueden agregar y modificar el contenido publicado de forma sencilla e instantánea; los marcadores sociales, plataformas en las que los participantes pueden buscar, almacenar y organizar contenido encontrado por otros usuarios; y los sitios de opinión, donde encontramos comentarios y valoraciones acerca de un producto o servicio, con el fin de ayudar y compartir su experiencia con otros usuarios.

A continuación, observamos los medios sociales multimedia, compuestos por aquellos sitios web en los que se utiliza y comparte contenido de forma audiovisual y digital. Engloban principalmente los espacios donde se difunden fotos, vídeos o música.

Por último, tenemos los juegos en línea o mundos virtuales, en los que se crean comunidades de jugadores que interactúan entre sí a través de personajes en línea o avatares dentro de un universo ficticio o inspirado en la realidad.

El creciente uso, popularidad y alcance de estos medios hace que cada vez más empresas los integren en su modelo de negocio, convirtiéndose en una de las herramientas más eficaces para que un producto o servicio llegue a los usuarios finales. Y no sólo para captar a consumidores potenciales, sino para aumentar la visibilidad de la marca y, consecuentemente, su reputación. Sin embargo, esto no resulta una tarea sencilla pues, pese a los grandes esfuerzos de las empresas por posicionar su propuesta de valor en el mercado, se enfrentan a competidores y costes que, parecen imperceptibles ante los clientes, pero suponen un gran gasto en la compañía.

Por lo tanto, si las empresas quieren incorporar y hacer un buen uso de estos medios en sus estrategias comerciales, deben comenzar por transformar y reestructurar los departamentos de marketing y comunicación, permitiendo la creación de nuevas oportunidades de negocio como la mejora en los mecanismos de relación con sus clientes. Tienen que hacer un uso correcto del potencial que ofrecen, y aprovechar que los clientes confían cada vez más en estos medios para sus compras, las compañías tendrán a su disposición un instrumento de marketing y promoción muy valioso (Yogesh y Yesha, 2014).

Los medios sociales permiten a las empresas realizar un análisis predictivo acerca de los comportamientos futuros de sus clientes, tendencias, estudios de mercado, desarrollo de nuevos productos o análisis de riesgos. Tal y como indica Bravo (2018), representan uno de los mayores recolectores de información social del mundo.

Por un lado, supone una herramienta de inteligencia competitiva, pues gracias a la información recolectada acerca de sus competidores y consumidores, se pueden desarrollar estrategias corporativas. Y por el otro, es un gran instrumento comunicativo, ya que invita al desarrollo de relaciones con los clientes, la promoción, publicidad y visibilidad de la marca, así como la atención al cliente (Sellés Revert, 2016).

Todo esto hace que las empresas se vean presionadas ante la necesidad de mantener su reputación online. Es por ello por lo que deben enfocarse en la comprensión de las influencias y el impacto generado por su uso, componentes que afectan el negocio (ILTADMIN, 2013). Se convierte el cliente entonces, en un elemento clave dentro del marketing y gestión de la reputación.

Esto se traduce en una nueva necesidad para las compañías, la de fomentar las relaciones e interactuar con sus clientes mediante la creación de comunidades. La idea general es que los consumidores se sientan escuchados, de manera que se genere un compromiso con la empresa y estos puedan compartir sus experiencias positivas con otros usuarios.

En definitiva, el impacto que han generado estas plataformas en las empresas es enorme, pues ha supuesto un cambio en la forma de conectar con los clientes, la adopción de nuevas estrategias comerciales, la creación de barreras de entrada para competidores, y el aumento de la demanda, entre otros.

2.1.1 Uso e impacto de estos medios en el contexto turístico

El uso y disponibilidad de estos medios sociales abarca un gran campo, incluido el sector turístico. Son ya muchas compañías turísticas las que aprovechan las ventajas y oportunidades que ofrecen estos mecanismos para promocionar destinos turísticos, instalaciones y servicios.

La figura del turista tradicional ha evolucionado a lo largo de los últimos años, exigiendo una oferta especializada y que se adapte a sus necesidades. Es decir, ha pasado de ser un turista observador a uno activo, donde los gustos y aficiones determinan su experiencia y la implicación en ciertas actividades (Machado Chaviano y Hernández Aro, 2008). Se convierte en su propio gestor turístico, es decir, que se mantiene informado y actualizado de manera

independiente, sin la necesidad de recurrir a agencias de viajes. Planifica su viaje con respecto a la información disponible en Internet, y según la oferta de los promotores turísticos (Suau Jiménez, 2012).

Estos cambios hacen que los gestores y compañías turísticas en general deban centrarse en desarrollar estrategias de marketing que satisfagan las necesidades de un viajero cada vez más informado y exigente, donde los medios sociales cobran especial importancia (Sánchez-Amboage et al., 2019). Deben enfocarse en la mejora de sus políticas para ofrecer productos y servicios únicos a través de una comercialización eficaz (Adarsh y Sreereshma, 2021), dando lugar a una nueva forma de competencia, y la adopción y desarrollo de nuevos modelos de negocio.

Tal y como explican en su artículo Sánchez-Amboage et al. (2019), las empresas turísticas buscan conseguir el mayor número de turistas, y la manera de hacerlo es mediante la planificación y creación de unas estrategias que reflejen la identidad e imagen del destino. Proyectar una imagen fiel de los servicios que se ofertan e implementar modelos de promoción y comunicación eficaces se convierten en tareas fundamentales para captar la atención del turista y convencerle de que escoja una empresa concreta en la programación de su viaje.

Normalmente, en esta planificación hay varios agentes implicados que van desde la reserva del vuelo, pasando por el alojamiento o el alquiler de coches, todo eso a través de Internet. Es por ello por lo que se recomienda a todas las compañías turísticas, familiarizarse con el uso y gran aportación que tienen los medios sociales en la industria del turismo (Icoz et al., 2018).

Por otro lado, el uso de estas plataformas ha supuesto una gran fuente de datos alternativa, ya que nos permiten seguir los movimientos y consumo de los turistas. Nos proporcionan información relevante y detallada sobre el comportamiento de estos, sus intereses, propósitos de la estancia o los lugares que visitan (Kovács et al., 2021). Todo este contenido no sólo puede ser utilizado por las empresas turísticas para conocer más a los clientes y sus preferencias hacia la marca, sino por los propios turistas. Estos últimos podrán acceder fácilmente a una gran cantidad de datos significativos de forma dinámica, para encontrar información sobre el destino u otras incertidumbres durante el proceso de decisión de compra (Silveira et al., 2021).

Dentro de la industria del turismo, nos centraremos en los alojamientos y cómo estos hacen uso de los medios sociales para promocionar sus establecimientos como destino turístico. Se enfocan en crear interés hacia sus instalaciones e introducir promociones, con el fin de

atraer la atención de los turistas para aumentar el reconocimiento y visibilidad de sus marcas. Uno de los casos más cercanos y actuales que encontramos es la promoción de estancias en las islas Canarias mediante la entrega de los "Bonos Turísticos Somos Afortunados". A través de un sorteo, los residentes canarios podrán disponer de una tarjeta monedero de 200 euros para emplearlo en los alojamientos turísticos inscritos al programa (Hola Islas Canarias, 2021). El objetivo de este proyecto fue incentivar el consumo regional para fomentar y recuperar el turismo de las islas tras los devastadores efectos de la COVID-19.

Como hemos visto, las empresas hoteleras tienen que ser capaces de adaptarse a los constantes cambios de comportamiento y procesos de decisión de compra de los consumidores y, por consiguiente, saber replantear las estrategias corporativas y de marketing, con el fin de optimizar y generar el máximo beneficio.

Por lo tanto, los objetivos que persiguen los alojamientos a través de la utilización de los medios sociales son poner a disposición del cliente potencial información pertinente y oportuna del hotel, y facilitar a los huéspedes la reserva del establecimiento (Tuominen, 2011).

De acuerdo con lo mencionado anteriormente, y atendiendo a los principales agentes que constituyen el sustento del sector turístico, se concluye que los gestores de hoteles deben comenzar por fortalecer la interacción y relación con sus clientes, tanto para solucionar los problemas puntuales como para informar acerca de sus servicios.

Estas relaciones con los clientes reducen la posibilidad de sustitución del servicio por otra empresa, y aseguran una rentabilidad a largo plazo. Se habla entonces de "lealtad" como elemento indispensable de marketing, ayudando a la creación de vínculos con los consumidores y la aparición de ventajas competitivas (Salvi et al., 2013).

Asimismo, debemos tener en cuenta la reputación corporativa como factor determinante, pues indica la percepción del cliente acerca de la capacidad de rendimiento del servicio y, consecuentemente, una mayor confianza por parte del consumidor. Esto ayuda no sólo a generar una ventaja competitiva frente a otras empresas sino la repetición de compra de los clientes.

En conclusión, los propietarios y gestores de hoteles, y de servicios turísticos en general, deben reconocer el papel que tienen estos medios para intensificar sus esfuerzos en

desarrollar y ampliar el uso de las actuales tecnologías interactivas con el fin de mejorar su posición en el mercado y maximizar sus beneficios.

2.2 Las valoraciones de los productos y servicios en Internet

El creciente uso y popularidad de los medios sociales ha hecho que la voz de los usuarios cobre cada vez más importancia y llegue a todas partes, originando una evolución del concepto de "boca en boca" hacia uno de "boca en boca electrónica" (Sarmiento Guede et al., 2017), o lo que se conoce comúnmente como *eWOM*. Esto es, la información que da un consumidor acerca de un producto o servicio a través de los medios sociales y que va dirigida especialmente a otros consumidores potenciales (Adarsh y Sreereshma, 2021).

Tal y como vimos en el apartado anterior, estas valoraciones las podemos encontrar en los sitios web de opinión tales como TripAdvisor, utilizado en el desarrollo del actual trabajo. Estos se identifican dentro de los medios sociales de colaboración, en el que los usuarios son libres de aportar y expresar una opinión acerca de un producto o servicio, con el fin de ayudar a otros posibles consumidores.

Una consecuencia directa del impacto que ha generado la utilización de estos medios es la desconfianza de los usuarios ante la información proporcionada por las marcas y el incremento en la fiabilidad de los comentarios de otros consumidores durante el proceso de decisión de compra (Paús y Macchia, 2014).

Dentro de este proceso de decisión de compra (ilustración 1), las opiniones juegan un papel muy importante, pues contribuyen a generar una imagen positiva de la marca y a que otros consumidores también apuesten por ella. Se puede entender como un proceso de retroalimentación, donde los usuarios generan contenido a modo de información, recomendaciones y experiencias para posibles clientes potenciales (Yuzdepski, s. f.).

Durante las primera etapas de información y evaluación de alternativas, los comentarios tienen mucha importancia porque son la primera imagen que nos formamos de la empresa, si un comentario nos llama la atención, es cuando decidimos investigar acerca de ella y los productos o servicios que ofrecen. Asimismo, nos interesa saber la reputación que tiene esa compañía o la valoración que tienen los bienes que producen, adquiriendo otra vez importancia las reseñas de los consumidores.

Ilustración 1. Proceso de Decisión de Compra

Fuente: Máñez (2019)

Por otro lado, estas valoraciones también cobran fuerza en la etapa de poscompra, pues un usuario puede influir en el proceso de compra futura de otro cliente potencial, generando y alimentando a la reputación de la marca mediante la comunicación de su experiencia. Esta fase del proceso es crucial y muy significativa para los proveedores, pues afecta a las ventas futuras y la opinión que se refleje sobre los productos.

A lo largo de esta última etapa es muy importante la respuesta que las empresas dan a esas reseñas, con el fin de fortalecer la relación con sus clientes y conservar su fidelidad. Las respuestas se pueden dar tanto a una valoración negativa, en donde se presta atención al problema y se establece una solución; como a una valoración positiva, para agradecer y animar a los clientes a comprometerse con la marca (Chan, 2017).

Está claro que la generación de valoraciones positivas ante un producto o servicio puede mejorar la reputación y fiabilidad de una marca. Cuantos más usuarios comenten la satisfacción obtenida ante la propuesta de valor de una empresa, más popularidad ganará esta, traduciéndose en un incremento en ventas y beneficios.

Otra de las principales estrategias que pueden adoptar las empresas es el aumento del conocimiento de la marca, de manera que se incremente su visibilidad y ayude a los consumidores a dar con ella. En este caso, se puede recurrir a un factor que está en auge actualmente: los *influencers*. Estos son personajes populares en medios sociales, seleccionados para promocionar contenido a sus seguidores, es decir, a los posibles consumidores de la firma. Un informe reciente afirma que el 94% de los profesionales del marketing que han recurrido a esta forma de promoción las consideran eficaces, por su

contacto más auténtico y directo con los clientes potenciales (Lou y Yuan, 2019). El objetivo que persiguen las empresas que utilizan este medio es proporcionar el producto o servicio al *influencer*, normalmente de forma gratuita, para que este lo valore y recomiende a sus seguidores con el fin de que se conviertan en consumidores potenciales y futuros clientes de la marca.

De acuerdo con lo descrito previamente, podemos concluir que las reseñas juegan un papel fundamental en la visión e imagen que se generan el resto de los usuarios acerca de la empresa, ya sea sobre su marca o gestión interna. Por ello, es esencial que las compañías mitiguen este impacto buscando siempre la satisfacción, compromiso y fidelidad de los clientes.

2.2.1 Las valoraciones de los servicios en el ámbito turístico y su impacto

Como ya hemos visto, los consumidores ya no solo se guían por los anuncios que hacen las propias empresas turísticas, sino lo que otros usuarios opinan sobre su negocio (Adarsh y Sreereshma, 2021). Es por ello por lo que los gestores turísticos deben hacer un esfuerzo en recopilar esa información y utilizarla para su beneficio.

Este contenido proporcionado por los turistas supone de un gran valor para la comprensión de las necesidades, mejoras y preferencias de las infraestructuras y servicios de los establecimientos dedicados al turismo, resultando en la modificación de muchas decisiones de gestión (Miguéns et al., 2008). Entender los patrones de viaje y comportamientos de los turistas, ayudará a las empresas a identificar destinos claves y mejorar las estrategias actuales (Sugiharti et al., 2021).

A la hora de planificar un viaje y, ante la gran variedad existente de alojamientos turísticos en Internet, los usuarios valoran positivamente que estos dispongan de comentarios que faciliten su búsqueda, y ayuden a filtrar características durante el proceso de decisión de compra. Es más, un estudio realizado por la Universidad de Ciencias Aplicadas de Worms reveló que el 46% de los usuarios señalaban los sitios de reseñas como la fuente de información principal a la hora de elegir el alojamiento (Gonzalo, 2014b).

Por lo tanto, es labor de los alojamientos turísticos, en este caso, hacer todo lo posible para que su establecimiento sea percibido como un excelente destino vacacional. Consiste en mostrar y demostrar una imagen fiel, auténtica y atractiva del establecimiento y sus servicios con el fin de atraer el máximo de turistas posible.

La mayoría de estas valoraciones, en el caso de los hoteles, se encuentran en sitios de opinión como TripAdvisor, una de las plataformas de viajes más grandes del mundo donde los usuarios son los principales protagonistas, pues son los que proporcionan la mayor parte del contenido: las reseñas.

A través de este sitio web, los clientes podrán compartir información más detallada de los servicios y, de alguna forma, se convertirían en "embajadores de la marca" de manera natural, recomendando y sabiendo que su opinión está siendo escuchada por otros (eTurboNews, 2020). Motivo por el que muchos turistas recurren a esta plataforma a la hora de planificar su viaje, para leer y saber lo que piensan otros usuarios sobre la calidad del servicio ofertado en el posible alojamiento elegido.

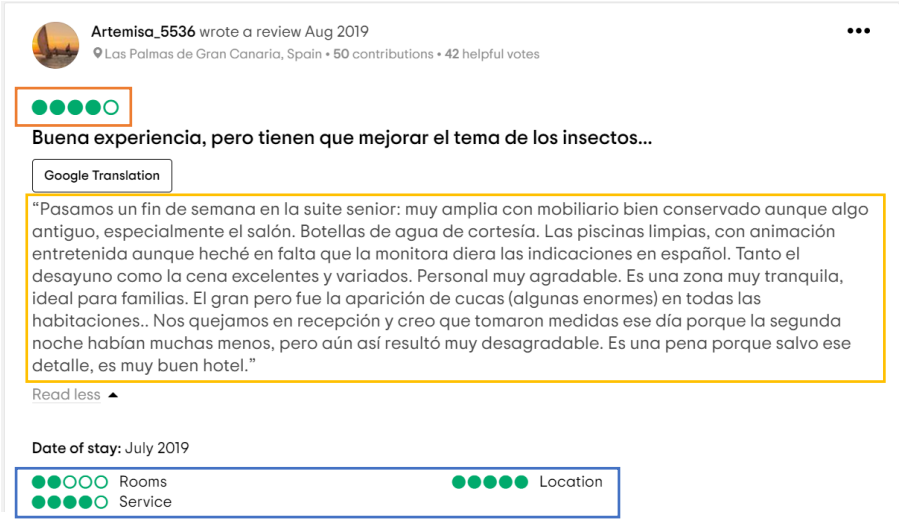
Esta afirmación se sustenta en un estudio realizado por *PhoCusWright*, líder mundial en investigación del sector de los viajes, en el que 77% de los participantes consultaban la plataforma TripAdvisor antes de elegir un hotel. Asimismo, el 80% de los usuarios afirmó leer entre seis y doce reseñas antes de reservar cualquier alojamiento turístico (Gonzalo, 2014a).

Esta misma investigación reveló que el 83% de los turistas, tras consultar las opiniones de TripAdvisor, escogieron el hotel correcto, y el 80% disfrutaron de un mejor viaje (Tripadvisor Insights, 2014). Es decir, que, en general, aunque las opiniones encontradas en esta plataforma puedan realizarse de forma anónima y no requieran haber estado físicamente en un alojamiento turístico para opinar sobre él, son fiables; gracias también al sistema de antifraude que tiene la plataforma. Tal y como explica en su página web, TripAdvisor se compromete a que el contenido mostrado sea coherente y refleje de forma auténtica la experiencia que han tenido verdaderamente los usuarios; siendo esta una de las razones por las que hemos escogido este sitio web para la implementación y desarrollo del presente proyecto.

A continuación, en la ilustración 2, observamos un ejemplo de comentario que podemos encontrar en esta plataforma. Como vemos, se compone de varios elementos: en primer lugar, del nombre del usuario que realiza la valoración, la fecha en la que se realiza y de dónde proviene; seguidamente encontramos la puntuación general, representada con cinco círculos verdes, que el cliente le otorga al alojamiento, en este caso, al Hotel Cordial Mogán Playa. A continuación, apreciamos el título y la descripción de la reseña;

terminando con un sistema de valoración individual para algunos de los aspectos más importantes, característica no encontrada en otras plataformas de viajes.

Ilustración 2. Opinión de TripAdvisor



Fuente: Elaboración propia a partir de TripAdvisor

En este caso, el hotel mencionado puede hacer uso de esta opinión para, como menciona el usuario, mejorar su experiencia mediante la fumigación de insectos en las habitaciones, de manera que no le vuelva a ocurrir a futuros clientes. Es tan importante el comentario del turista como la respuesta que le dan los gestores hoteleros, con el fin de que, tras su estancia, el viajero sienta que sus recomendaciones y quejas están siendo escuchadas. Será percibido positivamente por los consumidores, pues les aporta confianza que los hoteles dediquen su tiempo en responder y atender sus necesidades (Xie et al., 2016).

Las respuestas generadas por los gestores del marketing afectan la manera en la que los consumidores perciben la marca y están dispuestos a comprar. También, daría lugar a la generación de más reseñas, aumentando la relación entre el volumen de estas y el rendimiento del hotel. Esta respuesta no sólo mejorará el vínculo con los clientes, sino que lo diferenciará de sus competidores directos, pues genera una comunicación bidireccional entre ambos, que sirve de referencia a los futuros clientes potenciales (Xie et al., 2016).

Un estudio realizado sobre el impacto que generan las reseñas de clientes en las ventas de habitaciones de hotel, indica que un incremento de un 10% en las evaluaciones de las valoraciones suponía, más adelante, un incremento del 4,4% en las ventas (Ye et al., 2009). Por lo que, si lo juntamos con lo explicado anteriormente, concluimos que la implicación de un hotel con sus clientes y la búsqueda de su satisfacción da lugar a la generación de

más valoraciones por parte de los turistas y estas, a su vez, un aumento significativo en las ventas.

Por otro lado, cabe mencionar que TripAdvisor dispone de otra forma de interacción con los usuarios aparte de las mencionadas valoraciones: los foros de discusión. Estos permiten a los consumidores compartir información, consejos y experiencias con la comunidad, y suponen casi 2800 nuevos temas de discusión al día (Mkono y Tribe, 2017). Esto implica una gran herramienta para estimular la interactividad entre los usuarios mediante recomendaciones y consejos no sólo del alojamiento en el que el cliente se vaya a hospedar, sino de atracciones turísticas cercanas, restaurantes, actividades de ocio…

Para concluir, nos damos cuenta de que las valoraciones pueden determinar el rendimiento de los proveedores de servicios y afectar al valor percibido, su imagen y reputación; influyendo también en las decisiones de los consumidores, como la probabilidad de reservar, la fidelidad o la intención de recomendar (Taecharungroj y Mathayomchan, 2019). En definitiva, si somos una empresa dedicada al turismo y queremos prosperar en el negocio, debemos aprovechar las ventajas que nos aportan los medios sociales, y atender a las necesidades de los clientes de una forma efectiva para garantizar su satisfacción y hacer que otros consumidores potenciales reconozcan e inviertan en tu marca.

2.3 Métodos y técnicas de análisis

En este apartado se explicará la teoría de las distintas técnicas de Inteligencia Artificial llevadas a cabo en el proyecto. Comenzaremos estudiando el análisis de texto mediante el Procesamiento de Lenguaje Natural, y el Análisis de Sentimientos. Se expondrán también las distintas aplicaciones que este último tiene, así como los retos o limitaciones con los que se encuentra hoy en día. Terminaremos explicando una de las ramas más complicadas y específicas de este análisis: el análisis de sentimientos a nivel de aspecto. Además, se presentará el concepto de redes neuronales y sus diferentes tipologías, junto con los modelos arquitectónicos *Transformer* y BERT.

2.3.1 Natural Language Processing (NLP)

El Procesamiento de Lenguaje Natural es un subcampo de la inteligencia artificial que se encarga de las interacciones entre humanos y máquinas. En un principio, cuando una persona generaba un mensaje, para que éste pudiera ser comprendido por otro individuo, se hacía necesario una tercera que ejerciera de traductor, si el idioma no era común entre

los extremos. El NLP ha hecho que este intermediario sea sustituido por máquinas que, no sólo traducen el mensaje, sino que analizan y obtienen información de forma útil e inteligente, a partir de los datos aportados por el emisor (Patel, 2020a).

Ilustración 3. Evolución del NLP

Fuente: Patel (2020)

Actualmente, tal y como podemos observar en la ilustración 3, constituye uno de los campos con mayor crecimiento tanto en lo que se refiere a la investigación como en lo relacionado con el crecimiento del volumen de negocio y desarrollo de nuevos productos y servicios en el mundo. Algunas de sus aplicaciones actuales son el reconocimiento del habla, la comprensión del lenguaje natural, los *chatbots*, la recomendación basada en el historial de búsqueda, la traducción automática o el análisis de sentimientos, del que hablaremos en el próximo apartado (Kasaraneni, 2020).

En el futuro se espera que estas máquinas sean más inteligentes, pudiendo aprender de la información disponible y aplicarla en el mundo real. Además, gracias a los avances en el NLP, los ordenadores serán capaces de recibir y dar información más relevante y útil (Patel, 2020a).

El Procesamiento de Lenguaje Natural se puede dividir a su vez en *Natural Language Understanding* (NLU) y *Natural Language Generation* (NLG). El primero, como su propio nombre en inglés indica, se esfuerza por comprender la importancia de los datos, entender su significado para, posteriormente, procesarlos. Para poder llevar a cabo este proceso, se utilizan diferentes técnicas y modelos que ayudan a comprender el contexto, la semántica, sintaxis, la intención o el sentimiento del texto. Su objetivo reside en encontrar la intención detrás de esos datos (Kaur, 2021). Uno de los mayores desafíos a los que se enfrenta es conocer la estructura y reglas del lenguaje, pues comprender el lenguaje natural actual sigue siendo un reto complicado para la IA (Dialani, 2020), pero al que se dedican muchos esfuerzos y se están consiguiendo avances sorprendentes.

Por otro lado, el NLG se encarga de la reproducción y generación de lenguaje natural a partir de datos estructurados, de manera que sea entendible por los seres humanos (Kaur, 2021). Algunos ejemplos de esto son los ya mencionados *chatbots* o asistentes virtuales de voz, pues son capaces de generar lenguaje ante un estímulo.

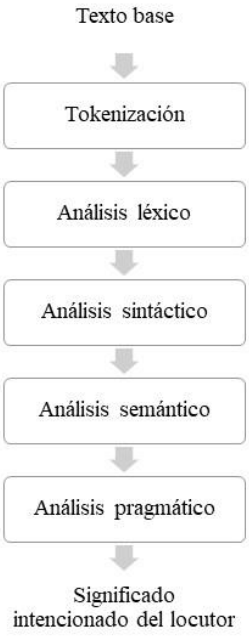
2.3.1.1 Natural Language Processing Pipeline

La transmisión de mensajes a la hora de comunicarnos se realiza mediante palabras, es decir, datos no estructurados. Estos suponen la mayor parte de la información circulante del mundo. Es tarea del Procesamiento de Lenguaje Natural, organizar y resolver los problemas derivados de estos grandes volúmenes de datos textuales. Para facilitar el proceso, se divide el problema en componentes más pequeños, de manera que se resuelva cada parte por separado, con ayuda del aprendizaje automático (Patel, 2020b).

Tradicionalmente, el análisis del lenguaje ha estado compuesto por tres etapas: sintaxis, semántica y pragmática. Se comienza por un análisis para determinar la estructura y orden de un texto, para pasar posteriormente a uno semántico, y acabar con una determinación del significado y contexto de los datos (Indurkhya y Damerau, 2010).

Sin embargo, la necesidad actual de tratar los datos lingüísticos reales hace que el proceso tenga una descomposición más detallada, tal y como observamos en la ilustración 4.

Ilustración 4. NLP Pipeline



Fuente: Elaboración propia a partir de Indurkhya y Damerau (2010)

1. Procesamiento del texto

En este primer paso, definimos la segmentación de frases y el proceso de tokenización como fases iniciales imprescindibles, puesto que los datos entrantes no están compuestos por frases cortas, delimitadas y bien formadas. Al dividir el texto en frases, se simplifica el proceso y se obtienen resultados más precisos. Estas oraciones, a su vez, se dividirán en componentes más pequeños llamados *tokens*, que pueden ser palabras, números o signos que puntuación que obtenemos mediante el proceso de tokenización (Tavva, 2021).

2. Análisis léxico

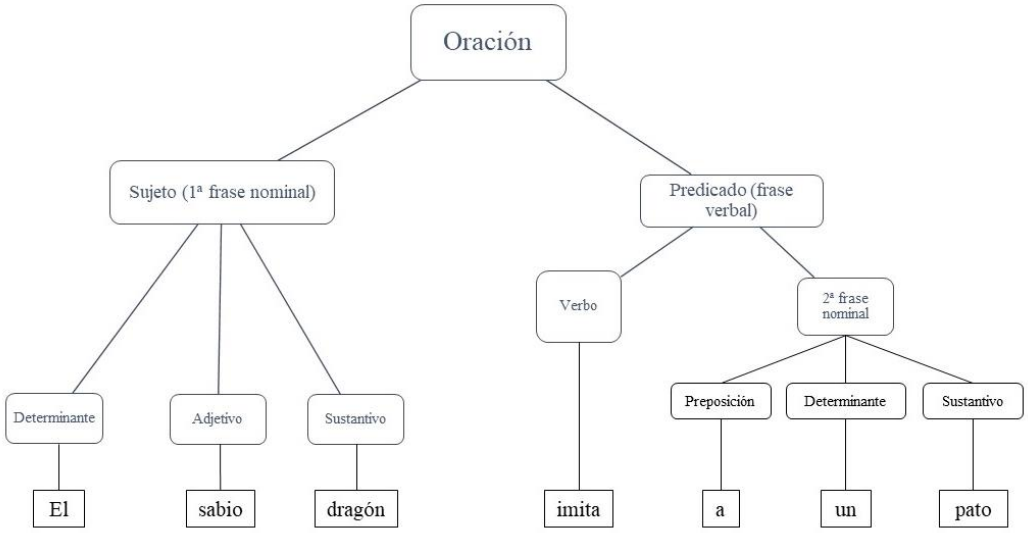
Dentro de los bloques de texto, las unidades más importantes de información son las palabras, por lo que se requieren de técnicas y mecanismos que analicen el texto a ese nivel, averiguando la raíz o lema de cada una, y registrando sus propiedades estructurales. En este análisis se pretende normalizar la palabra, sin tener en cuenta sus modulaciones, para poder utilizarla en aplicaciones como la clasificación de texto o recuperación de la información (Patel, 2020b). Los lemas de cada palabra se encuentran recogidos en un diccionario que nos permite relacionar cada variante morfológica con estos, además de incluir la información semántica y sintáctica (Indurkhya y Damerau, 2010).

Aunque el resultado obtenido será muy distinto a lo que teníamos inicialmente, este procedimiento pretende captar la esencia de lo que transmite el mensaje entrante, de manera que los datos conseguidos nos faciliten el trabajo posterior (Kasaraneni, 2020).

3. Análisis sintáctico

Comprende todas las técnicas básicas para analizar una cadena de palabras y determinar su estructura gramatical. Esto implica saber identificar el sujeto y predicado, así como el lugar de los sustantivos, verbos o pronombres en una oración (Mallamma y Hanumanthappa, 2014). Se centra en la combinación de palabras que forman una frase, donde cada una tiene una categoría diferente, al igual que reglas específicas de combinación. La siguiente figura muestra un ejemplo de este procedimiento, donde, dentro de cada frase nominal, se clasifican las palabras acorde a su gramática (González García et al., 2019).

Ilustración 5. Análisis sintáctico del NLP



Fuente: Elaboración propia a partir de González García et al. (2019)

La consecuencia del análisis es una estructura sintáctica jerárquica adecuada para el siguiente paso, que corresponde con la interpretación semántica.

4. Análisis semántico

Esta etapa se centra en la compresión del significado e interpretación de las palabras, los signos y estructuras de las frases. Tiene como objetivo estudiar el significado del lenguaje, es decir, la forma en la que las palabras y oraciones se refieren a los elementos de la naturaleza (Mallamma y Hanumanthappa, 2014).

El análisis no sólo se centra en buscar una explicación individual a las palabras, sino al significado de estas en la oración, y el de la frase en sí. De esta manera, se obtiene el enunciado de la frase independientemente del contexto.

5. Análisis pragmático y generación de lenguaje natural

Tras encontrar el significado de la frase, debemos estudiar su contexto, el uso y la comunicación que se hace de la lengua. Hace referencia a cómo se utilizan las oraciones en diferentes situaciones y su interpretación para conocer los objetivos e intenciones del emisor (González García et al., 2019). Estas estrategias van más allá de del significado literal de las palabras, apoyándose en los principios generales de la comunicación humana (Mahler et al., 2017).

Por último, debemos convertir ese pensamiento en lenguaje, es decir: identificar los objetivos del enunciado, planificar la forma de alcanzarlos mediante la evaluación de la situación y recursos disponibles, y generar los resultados en forma de texto (Indurkhya y Damerau, 2010).

2.3.2 Análisis de Sentimientos

Liu (2015) define el análisis de sentimiento como el estudio computacional de las opiniones, sentimientos, emociones y actitudes de las personas, cada vez más importante en los negocios y la sociedad.

Se trata de un subcampo del NLP perteneciente al análisis pragmático, que permite, automáticamente, sacar conclusiones sobre el estado de ánimo a partir de datos en formato de texto. Se puede realizar a tres niveles, con sus propios procedimientos y funciones: nivel de documento, de oración y de precisión o aspecto (Sütçü y Aytekin, 2019).

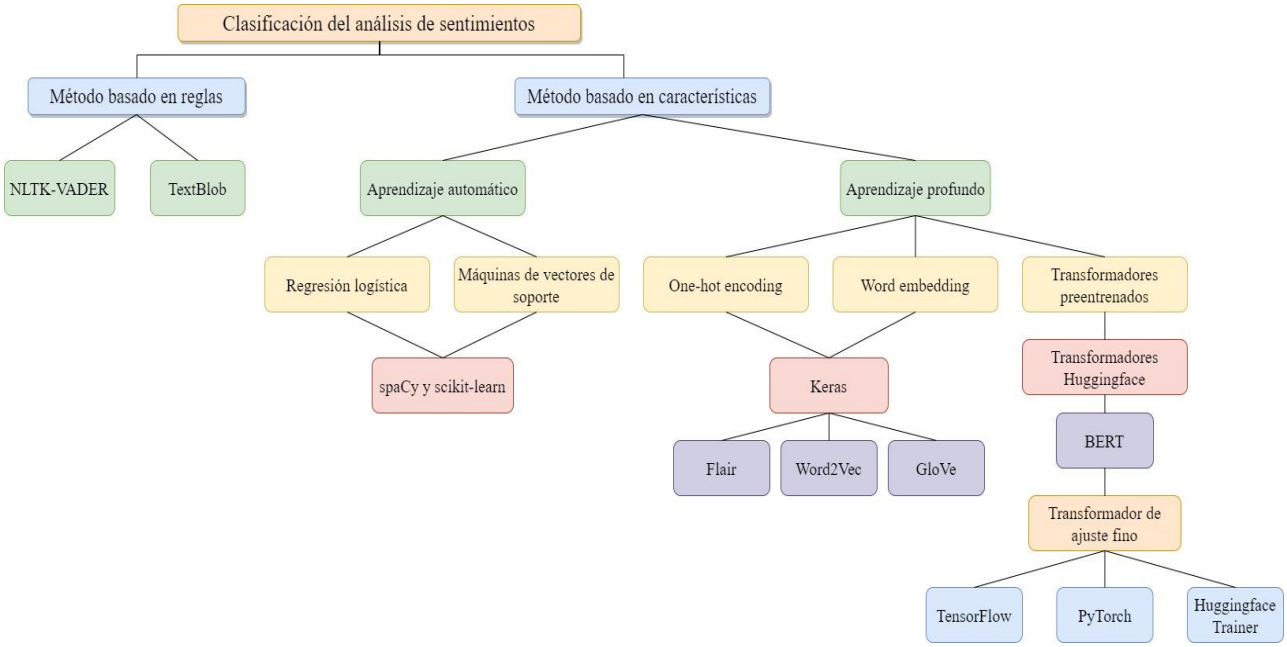
A nivel de documento, el análisis se hace de todo el texto y se genera un sentimiento general, clasificando el escrito analizado como positivo, negativo o neutro. En el siguiente nivel se diferencian dos aspectos. Por un lado, la subjetividad de la frase (objetiva o subjetiva), y por el otro, la clasificación sentimental de las frases subjetivas (positiva o negativa). Tenemos que hacer esta diferencia porque una frase subjetiva implica unos sentimientos, unas emociones. Se trata de un paso intermedio para descartar las frases que no expresan ninguna opinión y determinar si el sentimiento hacia una entidad es positivo o negativo. Por último, se analiza la oración a nivel de aspecto, donde se extraen las características del producto o servicio que se está valorando para, posteriormente, clasificar la opinión positiva o negativamente (Katrekar, 2019). Identifica directamente la opinión y su objetivo, permitiendo comprender mucho mejor el problema del análisis de sentimiento.

Hoy en día, y haciendo referencia a lo que mencionamos en los apartados anteriores, se crea cada vez más y más contenido generado por el usuario, permitiendo que estos últimos puedan expresar libremente su opinión, y que el análisis de estas emociones se vuelva más necesario a la hora de estudiar los medios sociales (Liu, 2015). Se utiliza en investigaciones de mercado de empresas u organizaciones para medir la eficacia de las actividades de marketing y la gestión de la reputación online (Sütçü y Aytekin, 2019).

El análisis de sentimiento se puede efectuar de dos maneras diferentes, tal y como vemos en la figura 6: con métodos basados en reglas, como los utilizados por las librerías de

Python *TextBlob* y *NLTK-VADER*, y con métodos de aprendizaje automático que se basan en características, que incluyen algoritmos de clasificación en una red neuronal de una sola capa, o profundos, que requieren múltiples capas (Naushan, 2020).

Ilustración 6. Clasificación del análisis de sentimientos



Fuente: Elaboración propia a partir de Naushan (2020)

2.3.2.1 Aplicaciones

El uso de aplicaciones de análisis de sentimientos se ha extendido a múltiples áreas, desde las finanzas, la biomedicina, los comercios minoristas, o los análisis de opinión y la política. Se trata de una herramienta fundamental y muy útil, que puede generar fuertes ventajas competitivas a la hora de analizar datos relevantes para las empresas.

En primer lugar, y haciendo referencia a lo que este proyecto engloba, se puede utilizar el análisis de sentimientos para monitorizar y conocer las opiniones de los consumidores ante un determinado producto o servicio, no sólo importante para las empresas sino para los propios clientes, que investigan y se preocupan por saber las valoraciones de un servicio antes de adquirirlo (Liu, 2015). Es una de las aplicaciones más prácticas y directas, ya que las compañías pueden conocer la opinión de sus clientes sin necesidad de realizar métodos tradicionales como las encuestas de satisfacción (Pauli, 2019). Además, con el estudio de estas valoraciones, las empresas podrán ofrecer un servicio mejor, crear productos más adaptados a sus necesidades o establecer nuevas estrategias de marketing para atraer a nuevos clientes. Los resultados que generan el análisis de sentimientos sobre estas

opiniones ayudarían a las empresas a conocer la percepción que tienen los consumidores sobre las tendencias actuales, y establecer relaciones más cercanas con ellos (Dang et al., 2020).

Por otro lado, como encontramos en muchos sitios web de reseñas, la valoración puede ir acompañada de una puntuación que, en muchas ocasiones, la opinión expresada no corresponde con esa calificación. Por lo que se podría utilizar el análisis de sentimientos para analizar el texto del usuario y corregirla (Pauli, 2019).

Algunos ejemplos de esto incluyen un estudio donde se aplicaron redes neuronales recursivas para analizar los sentimientos en las reseñas, con el fin de mejorar y validar las recomendaciones de restaurantes y películas de un sistema de recomendación basado en la nube (Preethi et al., 2017). Igualmente, se realizó un análisis de las influencias sociales en las reseñas de libros online, mostrando la relación existente entre las opiniones actuales y pasadas con la finalidad de comprender los comportamientos de contagio y las influencias entre los usuarios (Sakunkoo y Sakunkoo, 2009). Otro ejemplo es la tecnología *SumView*, un sistema de resumen de reseñas basado en la web, para extraer automáticamente las expresiones más representativas y las opiniones de los clientes sobre diversas características de los productos (Wang et al., 2013).

Aparte de la disponibilidad de un gran volumen de datos de opinión en los medios sociales, las opiniones y los sentimientos también tienen una gama muy amplia de aplicaciones, simplemente porque las opiniones son fundamentales para casi todas las actividades humanas (Liu, 2015).

Dentro del campo de la medicina también se ha empezado a utilizar este tipo de herramientas, por ejemplo, ofreciendo nuevos enfoques y proponiendo un léxico médico para apoyar a los expertos y a los pacientes en la variada metodología que se utiliza para describir síntomas y enfermedades (Satapathy et al., 2017).

Asimismo, un estudio llevado a cabo por el *New England Journal of Medicine* demostró que el 97% de los médicos estaban de acuerdo en que escuchar la voz del paciente era vital para mejorar su atención, por lo que se podría utilizar el análisis de sentimiento de audio para analizar sentimiento del hablante a partir de las señales de voz y así adquirir características representativas denominadas vectores de sentimiento de audio (ASV). Se emplearía el reconocimiento automático del habla (ASR) para convertirlo a una transcripción y, posteriormente, un análisis de texto (Repustate, s. f.).

Además de lo mencionado, otro estudio propone la implementación del análisis de sentimiento para identificar síntomas de depresión a través de los dispositivos móviles. Consistiría en analizar las emociones expresadas en las conversaciones de *WhatsApp*, de manera que se convierta en una manera más indirecta de llegar al paciente en lugar de realizar preguntas o formularios que puedan incomodarlo (León Martínez, 2019).

Otra de las áreas donde se puede ejecutar esta tecnología es la política, donde los directores de campaña pueden hacer un seguimiento de la opinión de los votantes sobre distintos temas y su relación con los discursos y acciones de los candidatos (D'Andrea et al., 2015).

Con referencia a lo recién mencionado, encontramos un experimento realizado en 2016 tras las elecciones de EE. UU., donde se examinaron millones de *tweets* que hacían referencia a los candidatos Trump y Clinton. Se analizaron comentarios de usuarios de todo el mundo para clasificarlos de manera positiva, negativa o neutra. El resultado fue que para ambos hubo más *tweets* negativos, pero la proporción entre ambos sentimientos resultó más beneficiosa para Trump (León Martínez, 2019).

Finalmente, otro de los grandes beneficios de la implementación de esta herramienta es prever cuál será la evolución financiera de una empresa según la información obtenida de blogs, webs y redes sociales (Pauli, 2019). Esto es lo que hace el sistema *The Stock Sonar*, permitiendo a los inversores obtener la esencia de miles de artículos cada día y ayudándoles a tomar decisiones de negociación oportunas e informadas (D'Andrea et al., 2015).

Es indiscutible la magnitud de las técnicas de análisis de sentimiento y todos los sectores que puede abarcar. Una técnica que está cobrando cada vez más importancia debido a los crecientes avances en la tecnología y el aumento del volumen de datos en la web. Sin embargo, esta herramienta puede presentar varios desafíos y retos que debemos abordar.

2.3.2.2 Retos y limitaciones

El análisis de sentimientos sigue siendo una de las tareas más difíciles del procesamiento de lenguaje natural, pues es una tecnología relativamente nueva que aún le queda mucho camino por recorrer.

Uno de los principales problemas es la aparición de una palabra que indica un sentimiento, en una frase no expresa ninguno. Esto es muy común en el uso de preguntas y condicionales. También nos podemos encontrar con el caso contrario: una frase que no

contiene ninguna palabra definida en el léxico de sentimientos que puede implicar un sentimiento positivo o negativo (Liu, 2015).

Otro de los retos a los que se enfrenta es la detección de oraciones implícitas. La oración explícita indica una opinión clara y subjetiva, mientras que una implícita es una opinión sugerida, que no aparece de forma directa. Por lo tanto, las primeras son más fáciles de clasificar y detectar (Panico, 2018).

Todas las expresiones se pronuncian en un contexto en concreto y detectarlo sigue siendo una labor muy complicada para este análisis si no se especifica de manera explícita. Esto conlleva a un problema de polaridad, donde los adjetivos son especialmente sensibles (MonkeyLearn, s. f.).

Por otro lado, es muy común encontrar textos en los que se hace uso del sarcasmo o la ironía, aspectos no identificables por esta metodología, ya que normalmente son clasificados de manera inexacta como un sentimiento neutral cuando realmente expresan una opinión negativa con palabras positivas (Shahnawaz y Astya, 2017).

Actualmente y cada vez más, es muy común encontrar opiniones negativas y positivas en una misma frase. Sin embargo, ya se han elaborado métodos para extraer frases comparativas de varias opiniones y generar un resumen con las oraciones contrastadas. Esto es lo que se conoce como el resumen comparativo de opiniones contrastadas (COS) (Panico, 2018).

Otra limitación es el lenguaje utilizado en la implementación y desarrollo de estas técnicas y métodos. La mayoría de las investigaciones sobre análisis de sentimientos se centran únicamente en textos en inglés, por lo que la mayoría de los recursos, librerías y herramientas se han desarrollado únicamente en ese idioma. Utilizar estos recursos e investigaciones en otros idiomas suele ser muy difícil y generan resultados imprecisos (Shahnawaz y Astya, 2017).

Por último, la generación de contenido en la web es tal que muchas veces nos encontramos con opiniones que no corresponden al producto o servicio ofertado. Esto es lo que se llama opiniones falsas, creadas con el fin de dañar la reputación de la marca y engañar al lector, con la posterior consecuencia de, a la hora de realizar un análisis de sentimientos, estas opiniones resulten inservibles (Chandni et al., 2015).

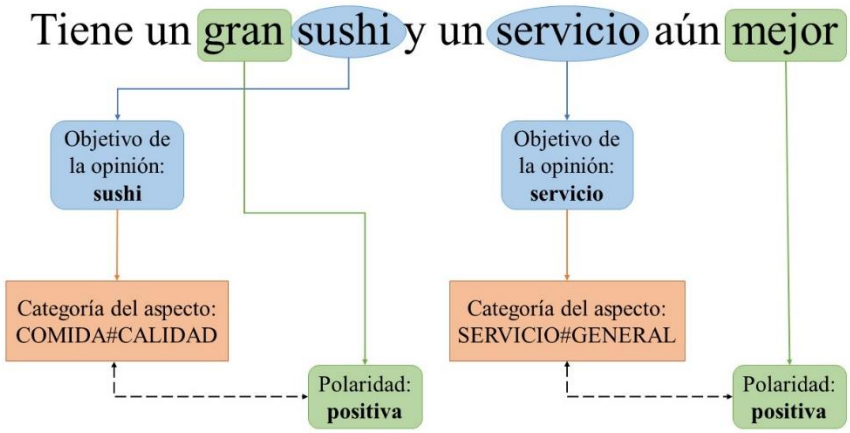
2.3.3 Aspect-based Sentiment Analysis (ABSA)

El análisis de sentimiento a nivel de aspecto o ABSA (Aspect-based Sentiment Analysis) es el grado más profundo y complejo del análisis de sentimientos, pues se basa en opiniones positivas y negativas sobre cada entidad y sus atributos, captando la esencia de las valoraciones y el sentimiento sobre estas (Liu, 2015).

ABSA se basa en identificar las características de una entidad para, posteriormente estimar la polaridad del sentimiento asociado a cada uno de esos aspectos. Además, permite a las empresas realizar un análisis más detallado sobre los comentarios de sus clientes para conocerlos mejor, y adaptar o crear productos que satisfagan sus necesidades (Pascual, 2019).

Este análisis se puede dividir en tres tareas principales: extracción de objetivos de opinión, detección de categorías de aspectos, y polaridad de sentimiento, tal y como vemos a continuación en la ilustración 7.

Ilustración 7. Ejemplo de las tareas del ABSA



Fuente: Elaboración propia a partir de Do et al. (2019)

Podríamos relacionar las dos primeras tareas, puesto que ambas se refieren a la extracción del aspecto de una entidad. Sin embargo, el objetivo de la opinión se diferencia de la categoría del aspecto en que el atributo se recoge de manera explícita, tal y como observamos en la imagen. Mientras que, en la categoría del aspecto, este puede aparecer tanto explícita como implícitamente en la oración (Do et al., 2019).

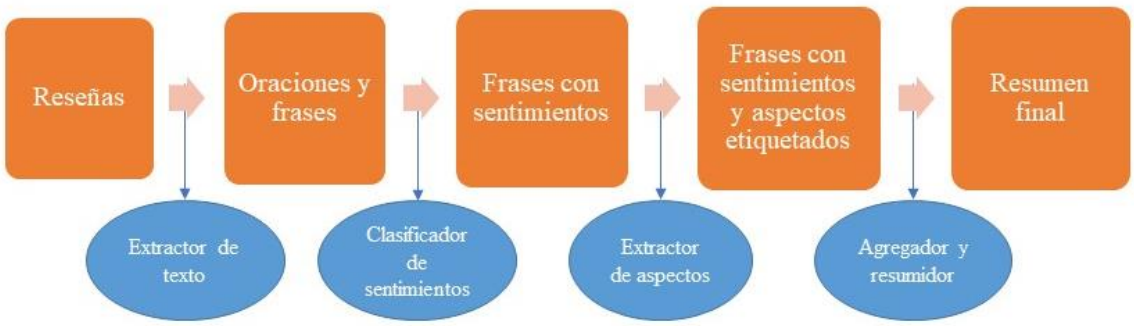
Para la extracción de los aspectos, se puede filtrar por reglas, como por ejemplo "aparece después de la palabra de sentimiento". En nuestro ejemplo, sería la palabra *service*, pues

viene precedida de una palabra recogida en el léxico de sentimientos: *better*. O, por el contrario, determinar manualmente todos los aspectos por adelantado y tratar de identificarlos en las reseñas. Por ejemplo, si tratamos con valoraciones de un restaurante, los aspectos podrían ser la comida, el servicio, la decoración del lugar, la localización… (Katrekar, 2019).

Una de las funcionalidades ampliamente utilizadas para detectar los aspectos es el *Topic Modelling*, donde se introduce el concepto de *tema* entre las variables de *documento* y *palabra*. De esta manera, el documento tendrá una mezcla de temas que a su vez se compone de palabras relevantes. Este enfoque es muy apropiado si se va a realizar un análisis a nivel de documento, pero puede ser complicado a niveles más profundos (Schouten y Frasincar, 2016). Además, en un estudio se observó que el uso de esta herramienta provocaba una incorrecta clasificación y evaluación de manera manual (Poria et al., 2016).

El siguiente paso consiste en analizar la polaridad o clasificación del sentimiento, que se puede realizar siguiendo los dos principios que vimos en el apartado anterior: técnicas basadas en aprendizaje automático o en reglas léxicas (Henríquez et al., 2017). Si nos centramos en este último veremos que hay diferentes léxicos de sentimientos disponibles que clasifican las palabras en positivas, negativas o neutras. Algunos ejemplos son *The General Inquirer*, *LIWC*, *MPQA Subjective Cues Lexicon*, *Bing Liu's Opinion Lexicon* o *SentiWordNet*. No obstante, estos léxicos en línea pueden ser incompletos al no contener suficientes palabras o no pertenecer al tema que se está tratando. En este caso, se debe construir un léxico manualmente con técnicas de aprendizaje automático, donde se estudian patrones para extraer frases objetivas y subjetivas (Katrekar, 2019).

Ilustración 8. Visión general del sistema ABSA



Fuente: Katrekar (2019)

Dentro del proceso que comprende el análisis de sentimientos a nivel de aspecto (ilustración 8), se extraen, en primer lugar, las diferentes oraciones que conforman el texto entrante. Y, a su vez, estas frases se dividirán en palabras y, posteriormente, en sus correspondientes lemas (Henríquez et al., 2017). El procedimiento más detallado de este paso lo encontramos en apartados anteriores donde nos referíamos al *pipeline* del procesamiento de lenguaje natural.

Seguidamente, se ejecuta un análisis de sentimientos a nivel de oración para determinar aquellas frases que guardan un sentimiento positivo, negativo o neutro y descartar aquellas irrelevantes, que no expresan ningún sentimiento.

Posteriormente, se obtiene el aspecto de aquellas oraciones que expresan un sentimiento y se establece su polaridad, para concluir con un resumen final de los aspectos extraídos con su correspondiente sentimiento.

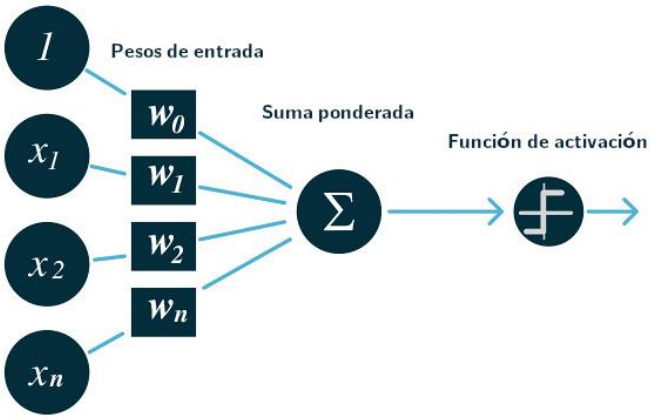
El análisis de sentimientos basados en aspectos trae consigo muchas ventajas puesto que permite a las empresas analizar grandes cantidades de datos en detalle, ahorrando dinero, tiempo y permitiendo concentrarse en otras tareas de gestión. Además, el análisis se realiza en tiempo real, ayudando a identificar los aspectos más relevantes y tomando medidas con el fin de satisfacer a sus consumidores. Todo esto implica la creación de una experiencia basada en el cliente, ofreciendo productos o servicios de una forma más rápida y personalizada (Pascual, 2019).

2.3.4 Redes neuronales

Las redes neuronales se basan en la biología, pues simulan neuronas interconectadas que procesan información como el cerebro humano. Siguiendo con esta idea, en 1943, McCulloh y Pitts crearon un modelo artificial de una neurona, que más tarde Rosenblatt llamaría perceptrón.

El perceptrón, tal y como observamos en la figura 9, está formado por varias entradas con sus respectivos pesos, y una función de activación que propaga la información según un umbral, generando un valor de salida. El modelo más simple de neuronas es cuando la función de activación devuelve un uno o un cero para indicar si la sumatoria de las entradas alcanzan el umbral de activación o no, respectivamente (Alias y Cassanelli, 2019).

Ilustración 9. Modelo de perceptrón

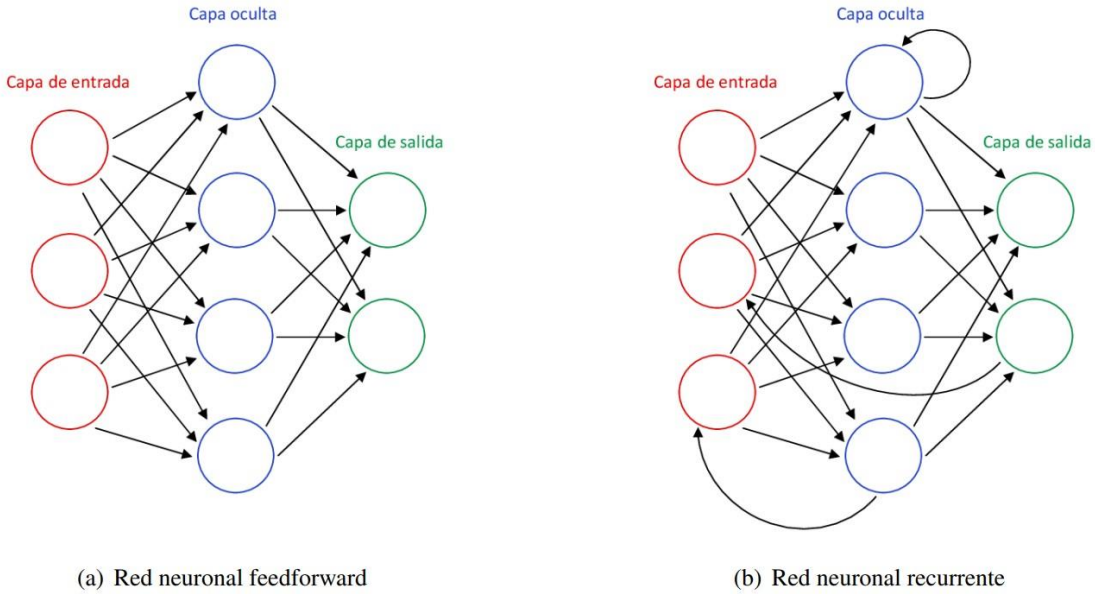


Fuente: Campos Soberanis (2020)

Las redes neuronales nacen de la interconexión de los perceptrones mediante sinapsis, organizándose en una estructura de capas. Estas redes contarán al menos con dos de ellas, una de entrada y otra de salida, pero podrán disponer capas ocultas, donde el valor de salida de una capa será el valor de entrada de la siguiente.

Diferenciamos entonces dos tipos de arquitecturas de redes (ilustración 10): las de propagación hacia adelante o *feedforward*, donde no existe la retroalimentación, sino que la salida de una capa es la entrada de la siguiente; y las redes recurrentes, donde se permiten las conexiones arbitrarias entre neuronas, permitiendo que la red tenga memoria (Aguilar Ibáñez, 2017).

Ilustración 10. Tipos de redes neuronales



(a) Red neuronal feedforward (b) Red neuronal recurrente

Fuente: Aguilar Ibáñez (2017)

Al principio los valores de los pesos se estiman de manera aleatoria y sin un criterio estipulado. Sin embargo, es aquí cuando entra el concepto de aprendizaje de máquina o entrenamiento mediante el cual se presentan ejemplos a la red en el que se conocen los resultados para comparar los valores de salida con las respuestas conocidas. A medida que este entrenamiento avanza, la precisión en los resultados se vuelve más firme, pudiendo aplicar las redes en casos futuros donde se desconoce el valor de salida final.

Cuando una red neuronal presenta más de tres capas, incluyendo las de entrada y salida, hablamos de Redes Neuronales Profundas, donde se pueden representar funciones de mayor complejidad ya que, cada capa se basa en los resultados de la anterior para refinar y optimizar la predicción o la clasificación (IBM Cloud Education, 2020). Dentro de estas redes diferenciamos las convolucionales, usadas en la clasificación de imágenes o vídeos para detectar y reconocer objetos; y las recurrentes, cuyas características veremos más en detalle a continuación, que se usan para el reconocimiento de lenguaje natural, mediante la utilización de datos secuenciales.

2.3.4.1 Redes recurrentes (RNN)

Tal y como vimos en el apartado anterior, las redes recurrentes pueden ser utilizadas para el procesamiento de lenguaje natural, objeto de este proyecto, pues tienen la capacidad de mantener un estado interno o memoria. Esto es realmente útil ya que el orden en el que aparecen las palabras es importante para analizar el contexto y significado final de los datos entrantes.

Estas redes se entrenan a partir de un conjunto grande y estructurado de ejemplos del uso de la lengua, un corpus lingüístico. De esta manera, los modelos de NLP podrán aprender los patrones necesarios para entender el lenguaje.

Un ejemplo del uso de estas redes lo podemos ver en el estudio titulado *Natural Language Generation, Paraphrasing and Summarization of User Reviews with Recurrent Neural Networks* (Generación de lenguaje natural, parafraseo y resumen de reseñas de usuarios con redes neuronales recurrentes), donde los autores exponen un modelo de red neuronal recurrente que puede generar nuevas frases y resúmenes de documentos (Tarasov, 2015).

Dentro de las redes recurrentes encontramos distintos subtipos:

- Redes recurrentes simples: los modelos hacen uso de una cierta memoria para aprender una secuencia temporal. No obstante, esta desaparece cuando se incrementa la secuencia de entrada.

- Redes LSTM: pensadas para solucionar los problemas donde la salida dependa de información antigua, pues introducen la estructura "celda de memoria". El contenido de ella está regulado por las puertas de entrada, salida y olvido.
- Redes GRU: creadas con el fin de solucionar el problema del gradiente. En lugar de utilizar una "celda de memoria", emplea estados ocultos y solamente dos puertas: una de reinicio y otra de actualización. Estas puertas mantienen un registro de la información relevante para la futura predicción. La primera decide cuánto contenido se debe transmitir al siguiente paso, y la segunda determina la información irrelevante y que se debe olvidar.

2.3.4.2 Modelo Transformer

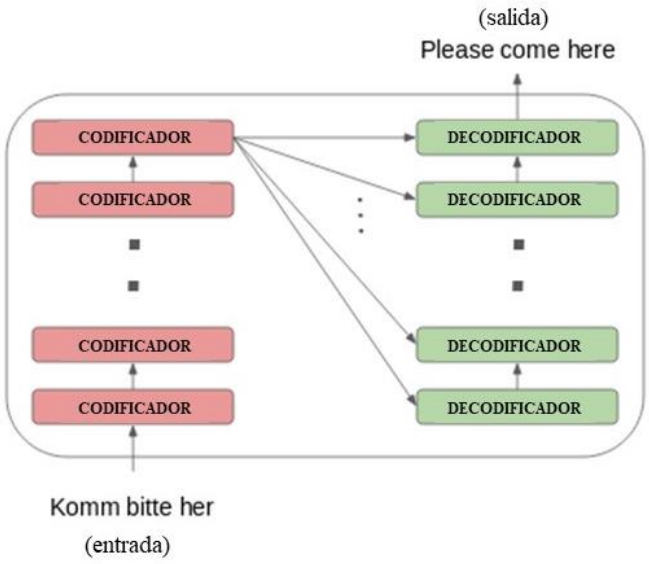
El modelo *Transformer* es un concepto arquitectónico muy actual, ubicuo y competitivo en lo referente a su desempeño. Se basa en el análisis de la secuencia de palabras que constituye la frase, utilizando el concepto de autoatención, donde se representa el contexto, asociando un peso a las relaciones entre las palabras. Esta autoatención, al contrario que las arquitecturas RNNs donde sólo se tenían en cuenta las palabras anteriores, genera el contexto distinguiendo tanto las palabras anteriores como las posteriores. Por ejemplo, en la frase "Llegué al banco después de cruzar el río", no identificaría la palabra "banco" como una entidad financiera, pues averigua su significado gracias a las palabras que se encuentran a continuación. Por tanto, es imprescindible conocer las relaciones y secuencias de palabras en las oraciones para que una máquina comprenda el lenguaje natural.

El concepto fue introducido por primera vez en el estudio *Attention is All You Need*, donde los autores propusieron una arquitectura de red basada en mecanismos de atención, tal y como explicamos anteriormente. Esta característica hace que la paralelización sea mucho mayor que en las RNNs y, por consiguiente, se reduzcan los tiempos de entrenamiento. Gracias a esta ventaja, se ha podido entrenar un volumen mayor de datos que con los sistemas anteriores (Vaswani et al., 2018).

La estructura del *Transformer* está compuesta por un codificador, que recoge la secuencia de los valores entrada y analiza su contexto; y por un decodificador, que genera la secuencia de salida a partir de ese contexto. Ambos se componen de módulos que se pueden apilar entre ellos, formados por capas multi-cabeza de atención y propagación hacia adelante. Las pilas tanto del codificador como del decodificador tienen el mismo número de unidades. En la

ilustración 11, vemos un sencillo ejemplo de traducción automática del alemán al inglés para entender su funcionamiento y la estructura general:

Ilustración 11. Traducción automática de forma simplificada con Modelo Transformer



Fuente: Joshi (2019)

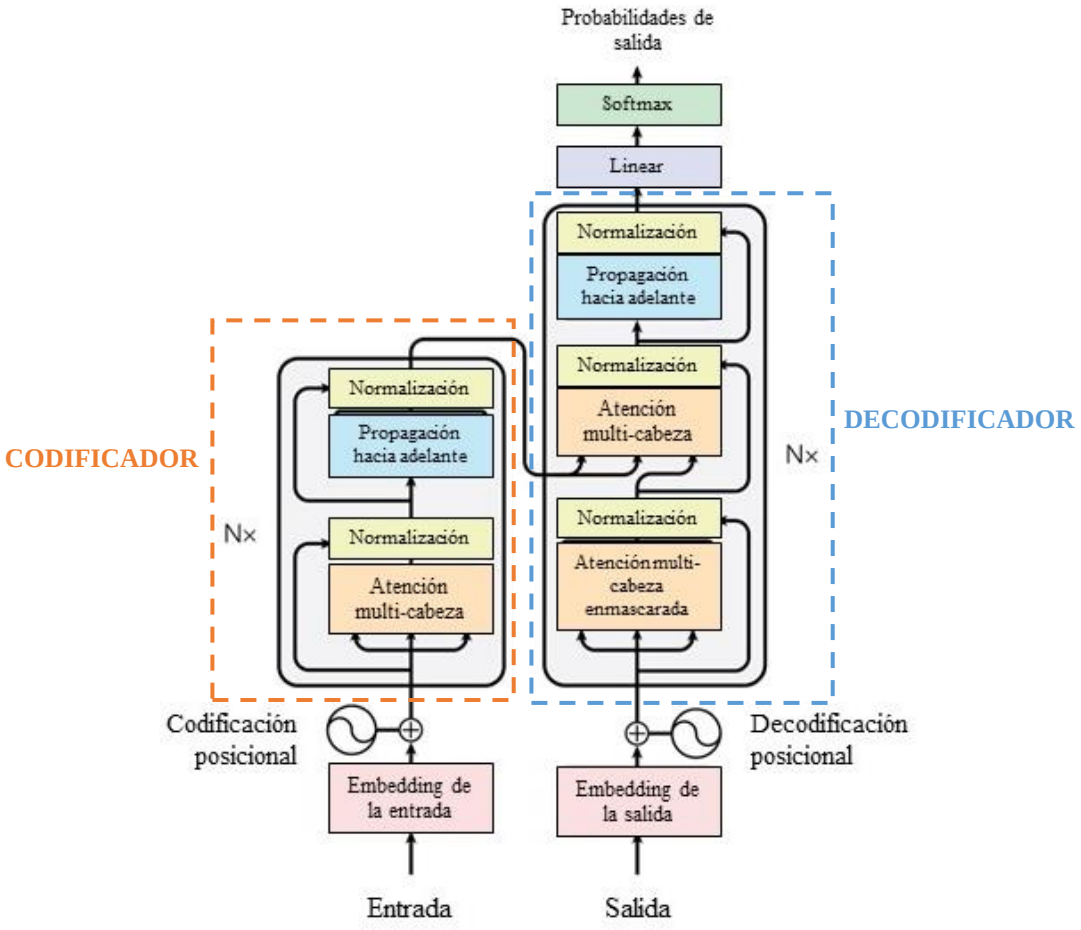
El codificador codifica la oración alemana en forma de vectores a través de mecanismos de atención y, por otro lado, el decodificador utiliza esa información para proporcionarnos la traducción al inglés. La salida de la última capa del codificador se traspasa a cada capa decodificadora, tal y como observamos en la imagen anterior.

Para que el codificador y decodificador puedan trabajar correctamente, es necesario realizar un *embedding* (empotramiento) del texto, es decir, la representación vectorial del conjunto de palabras. Además, se hacen necesarios los codificadores y decodificadores posicionales pues, al no trabajar con RNNs que recuerden la secuencia, necesitamos asignar a cada palabra una posición relativa para que sea entendible por el modelo.

La red se organiza por capas que se pueden ver de forma más clara en la ilustración 12, que representa la arquitectura del modelo *Transformer*.

Dentro de cada capa de la pila del codificador encontramos dos subcapas: un mecanismo de autoatención multi-cabeza y una de propagación hacia adelante; mientras que en el caso del decodificador encontramos tres: mecanismo de autoatención multi-cabeza enmascarado, otra autoatención sin máscara, y una de propagación hacia adelante.

Ilustración 12. Arquitectura del modelo Transformer



Fuente: Vaswani et al. (2019)

Cada una de estas subcapas es seguida por una de normalización, cuyo objetivo evitar que los datos caigan en la región de saturación de la función de activación, transformando la entrada en datos con media 0 y varianza 1.

La autoatención relaciona diferentes posiciones de una secuencia con el fin de calcular su representación para que el modelo pueda usar el contexto completo en la realización de la tarea. Esto se hace teniendo en cuenta tres valores: **Q** (representación vectorial de una palabra en la secuencia), **K** (todas las demás palabras de la secuencia) y **V** (valor vectorial de la palabra que se procesa en ese instante). El cálculo lo podemos encontrar a continuación:

Ecuación 1. Cálculo del mecanismo de atención

$$Atención(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Fuente: Vaswani et al. (2019)

1. Se calcula la puntuación o peso de cada palabra mediante el producto escalar de Q y K.
2. Se normaliza el resultado dividiendo por la raíz del tamaño de K.
3. Se emplea la función *Softmax* para obtener una distribución entre 0 y 1.
4. Se multiplica el valor resultante por el valor de la palabra actual para reducir la importancia de las palabras no relevantes y quedarnos con las que nos interesan.

Sin embargo, en la arquitectura del *Transformer* esta autoatención se calcula varias veces y de forma paralela e independiente. Es lo que se conoce como la capa de autoatención multi-cabeza que nombramos previamente, donde las salidas se concatenan y transforman linealmente. En la ecuación anterior se añade la variable **h**, que indica el número de "cabezas" que tiene el mecanismo. De esta forma se pueden aprender dependencias más complejas y sin añadir tiempo de entrenamiento.

A continuación, encontramos la capa *Feed Forward*, que se trata de una red neuronal que se aplica a cada vector de atención para transformarlos a una forma que sea entendible por la siguiente capa de codificación o decodificación.

Por otro lado, y refiriéndonos de la parte del decodificador, identificamos que la única diferencia con respecto al codificador es una tercera capa de atención multi-cabeza enmascarada. Esta subcapa se encarga de enmascarar ciertos valores para que no tengan ningún efecto en la actualización de parámetros. Por ejemplo, a la hora de realizar una traducción automática, se compararán los resultados esperados con la traducción real introducida en el decodificador. Sin embargo, debemos ocultar las palabras que se encuentren a continuación para que el mecanismo prediga las siguientes palabras por sí mismo y el aprendizaje tenga lugar (Ankit, 2020). Es decir, el modelo sólo tendrá en cuenta las palabras anteriores para aprender, las que se encuentran después debemos enmascararlas.

Finalmente, encontramos una capa lineal y de *Softmax* que se encargan de obtener las probabilidades de la siguientes palabras, y devolver aquella con el mayor valor como la que viene a continuación.

Este modelo condujo al desarrollo de sistemas preentrenados, como BERT (representaciones codificadoras bidireccionales de transformadores), que veremos en el próximo apartado, y GPT (transformador preentrenado generativo).

2.3.4.3 Modelo BERT

BERT o Representación de Codificador Bidireccional de Transformers es un modelo desarrollado por Google, basado en la arquitectura anterior de *Transformer* y utilizado en el preentrenamiento de aplicaciones de NLP.

Se presentó por primera vez en 2018, en el estudio titulado *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, donde los autores demostraban que un modelo de lenguaje entrenado bidireccionalmente ofrecía mejores resultados, pues se generaba una mejor comprensión y flujo del contexto (Devlin et al., 2019).

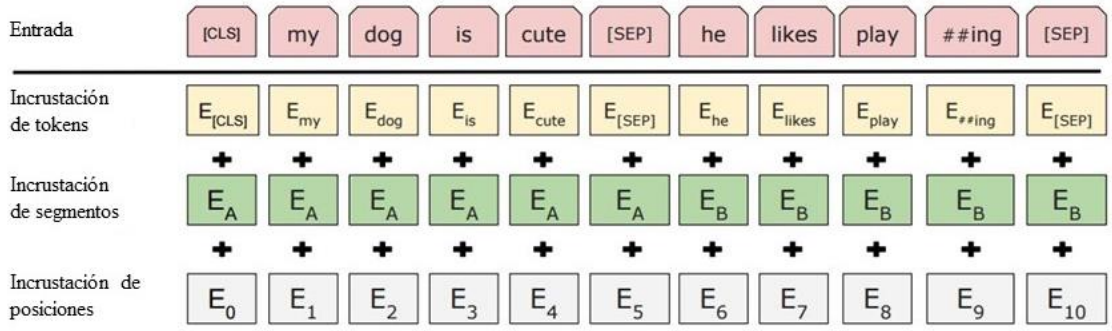
Una de las características más significativas y que diferencia a BERT de otros modelos es su bidireccionalidad, pues se examinan palabras tanto a la izquierda como a la derecha de cada término. Además, se trata de un sistema sin supervisión, es decir, que no requiere de un conjunto de datos finales para compararlos con los resultados obtenidos, sino que se concluyen directamente. Se trata de un modelo de lenguaje que fue preentrenado utilizando los corpus de texto *BooksCorpus* y Wikipedia, con un total de 3.300 millones de palabras.

Esta arquitectura sólo utiliza la parte del codificador del modelo *Transformer* que incluye cuatro etapas diferenciadas: la incrustación de palabras, la codificación de posición, el cálculo de autoatención, y la activación lineal.

En el preentrenamiento, nos basamos en dos tareas fundamentales: modelo de lenguaje enmascarado (MLM) y predicción de la oración siguiente (NSP). El objetivo de la primera de ellas es enmascarar algunos tokens de entrada aleatoriamente para predecir los valores originales de estas palabras, basándose exclusivamente en el contexto proporcionado por las que no lo están. Por otro lado, la siguiente tarea trata de averiguar si una oración es sucesora de otra devolviendo la etiqueta *IsNext* si efectivamente se trata de la siguiente frase, o *NotNext* si no es así, mediante la función *Softmax*. Se trata de una labor muy útil para tareas de comprensión de la relación entre oraciones, por ejemplo, en sistemas de respuesta a preguntas.

Dentro de la primera capa, y a diferencia del modelo *Transformer*, BERT incluye la codificación de segmento. Esto se hace necesario pues el modelo fusiona los pares de oraciones para elaborar el contexto. Las entradas están representadas por la suma de los vectores de inserción del token, segmento y posición, tal y como vemos en la ilustración 13 que se encuentra a continuación.

Ilustración 13. Representación de entrada del Modelo BERT



Fuente: Devlin et al. (2019)

1. Se inserta un token CLS al principio de la secuencia y uno SEP al final de cada frase.
2. Se agrega una incrustación de segmentos que indica qué token pertenece a cada oración.
3. Se agrega una incrustación posicional a cada token para indicar su posición en la secuencia.

Tras las actividades llevadas a cabo en el preentrenamiento, se inicia la fase de *fine-tuning* con los parámetros previamente entrenados. En esta etapa se mantienen los hiperparámetros del entrenamiento, salvo que alguno de ellos, especificados en la documentación BERT, requiera un ajuste. Dependiendo de nuestro objetivo y de la tarea NLP que queramos abordar, modificaremos el modelo añadiendo una capa de salida a la base BERT.

En la próxima ilustración encontramos algunas aplicaciones de NLP mediante el uso de este modelo. A continuación, describiremos los cuatro ejemplos que encontramos:

a) Clasificación de pares de oraciones

Encontramos ejemplos como MNLI (relación entre premisa e hipótesis), QQP (determinar si dos preguntas significan lo mismo) o STS-B (predecir la similitud de dos oraciones).

b) Clasificación basada en una oración

Algunos ejemplos son el análisis de sentimientos SST-2 de reseñas de películas o CoLA, tarea de clasificación para predecir si una frase en inglés es lingüísticamente "aceptable" o no.

c) Preguntas y respuestas

El sistema recibe una pregunta con respecto a una secuencia de texto y se requiere que marque la respuesta en la secuencia.

d) Reconocimiento de entidades nombradas (NER)

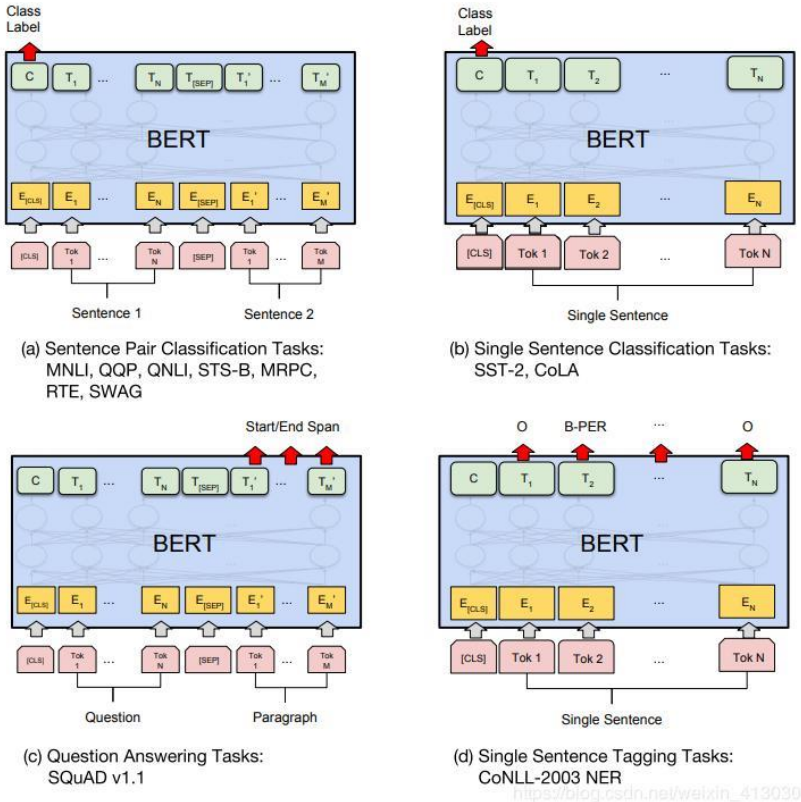
El software recibe una secuencia de texto y debe marcar los distintos tipos de entidades (Persona, Organización, Fecha, etc.) que aparecen en el texto.

Como hemos podido comprobar, los dos primeros ejemplos (a y b) son tareas a nivel de secuencia, mientras que los dos siguientes (c y d) son a nivel de token.

Actualmente, Google está utilizando este modelo para mejorar sus motores de búsqueda y procesar de manera más correcta el contexto de una oración. Por ejemplo, si en el buscador tecleamos la frase "2019 viajero brasileño a EE. UU. necesita visa" el modelo anterior no tenía en cuenta la preposición "a", y mostraba resultados sobre ciudadanos estadounidenses que viajaban a Brasil. Hoy en día el modelo es capaz de entender el contexto y ofrecer una información más apropiada al tener en cuenta la importancia de esa palabra en la oración (BBC News Mundo, 2019).

BERT es, sin lugar a duda, un claro ejemplo de la rápida evolución en el aprendizaje automático para el procesamiento de lenguaje natural, pudiendo implementarlo en diferentes aplicaciones (ilustración 14).

Ilustración 14. Ejemplos de tareas NLP con el Modelo BERT



Fuente: Devlin et al. (2019)

2.4 CRISP-DM (Cross-Industry Standard Process for Data Mining)

Para implementar cualquier proyecto tecnológico se hace necesario seguir y disponer de una metodología que ayude a organizar y seleccionar aquellas técnicas y métodos que se adecuen a las tareas pertenecientes al trabajo con el fin de obtener unos buenos resultados. En este caso, como se trata de un proyecto de minería de datos y estos últimos engloban un volumen grande de información, precisamos de un modelo que los gestione y analice de forma fiable. Dentro de la comunidad científica, CRISP-DM es una de las más populares, pues es de libre distribución, lo que significa que está en constante desarrollo, y es independiente de la herramienta que se utilice.

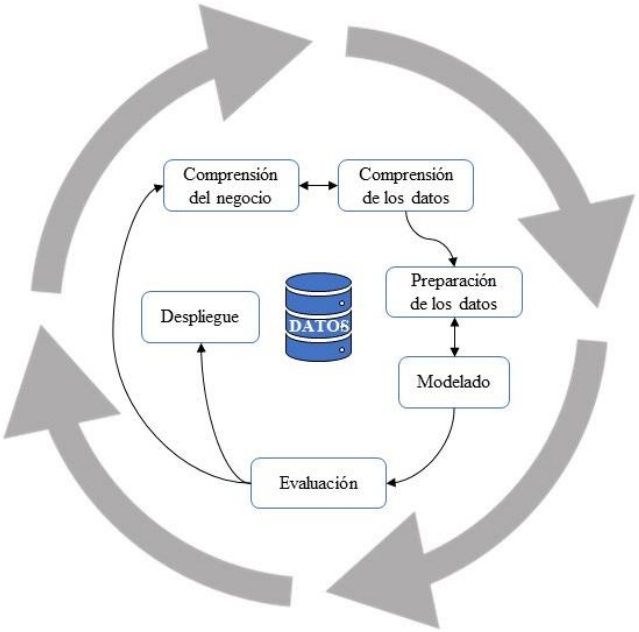
Esta metodología es una de las más utilizadas en la minería de datos, ya que enfoca sus resultados hacia el contexto del negocio. Se puso en marcha por primera vez en 1997, bajo un programa de financiación de la Unión Europea, y dirigido por cinco grandes empresas: SPSS, Teradata, Daimler, MCR y OHRA. La primera versión de esta metodología se presentó dos años más tarde en Bruselas junto con la publicación de una guía de minería de datos.

En las últimas dos décadas, el uso de las redes sociales y la capacidad de almacenar e intercambiar datos han aumentado considerablemente las oportunidades de extraer información mediante proyectos de minería de datos (Martinez-Plumed et al., 2019). Es por ello por lo que CRISP-DM se volvió necesario y cada vez más popular, debido a su capacidad de adaptación a las necesidades concretas de un negocio.

A continuación, veremos las seis etapas en las que se divide el ciclo de vida de la minería de datos (ilustración 15):

Etapa 1. **Comprensión del negocio:** Esta primera etapa es crucial para comprender los objetivos y exigencias del proyecto de *data mining* y alinearlos con los del negocio. Sin la comprensión de estos objetivos, ningún algoritmo obtendrá resultados fiables. Sus actividades principales comprenden: la determinación de los objetivos comerciales con el fin de identificar el problema inicial, la necesidad de utilizar la minería de datos y concretar los criterios de éxito; la evaluación de la situación actual, definiendo los requisitos del problema en términos del negocio y de *data mining*; definir los objetivos a nivel de minería de datos, puntualizando las metas a lograr tras la propuesta de solución; y, por último, la producción de un plan de proyecto, cuya base la forman las cuestiones y objetivos planteados hasta ahora, con la descripción de los pasos a seguir y sus correspondientes técnicas.

Ilustración 15. Ciclo de vida minería de datos



Fuente: Elaboración propia a partir de IBM (s.f.)

Etapa 2. **Comprensión de los datos:** En este segundo paso se busca recolectar y estudiar los datos disponibles con el objetivo de familiarizarse con ellos e identificar la calidad de los mismos. Se reconocen cuatro tareas dentro de esta fase: la recopilación de datos, ya sean existentes, adquiridos o adicionales; la descripción de los datos (cantidad, tipo, esquemas de codificación) ; la exploración de datos con el fin de encontrar una estructura general mediante pruebas estadísticas (tablas, gráficos y otras herramientas de visualización); y la verificación de calidad, incluyendo la identificación de datos erróneos, perdidos o incoherencias en la codificación para asegurar la completitud y corrección de los datos.

Etapa 3. **Preparación de los datos:** El objetivo final de este paso es obtener los datos finales con los que se trabajará y aplicarán los modelos. Guarda relación con la próxima etapa de modelado, pues según la técnica elegida en esta última, los datos se procesarán de una forma u otra.

Es una de las etapas que más tiempo y esfuerzo del proyecto consumiendo entre el 50-70% (IBM, 2020). puesto que incluye labores como: la selección de un subconjunto de los datos adquiridos en la etapa anterior mediante el uso de criterios preestablecidos; la limpieza de datos mediante diferentes técnicas; la creación de indicadores para fomentar la capacidad predictiva de los datos; y la transformación de los datos, cambiando su formato o estructura con la finalidad de prepararlos para la siguiente fase.

Etapa 4. **Modelado:** En esta fase, todos los datos obtenidos de las fases anteriores se incorporan a las herramientas analíticas y se empiezan a generar soluciones al problema planteado inicialmente. Aquí se deciden qué técnicas se van a emplear para un proyecto de *data mining* específico según unos criterios: adecuación al tipo de datos y capacidad para responder al problema planteado.

Las actividades incluidas en esta etapa son: la selección de técnica del modelado, de acuerdo con el problema a resolver, los datos y herramientas disponibles; la selección de datos de prueba si se requiere separarlos en datos de prueba y evaluación; y la obtención del modelo. En esta última tarea, es recomendable experimentar con varios modelos antes de llegar a conclusiones definitivas y compararlos entre sí.

Etapa 5. **Evaluación:** Comprende estudiar el acercamiento del modelo a los objetivos del negocio. Busca evaluar la calidad del modelo en base a determinadas métricas estadísticas y analizando los resultados. Si estos últimos son desfavorables o se presentan anomalías, debemos volver a la fase inicial de comprensión del negocio para replantear los objetivos. Si, por el contrario, se obtienen los resultados previstos, continuaremos hacia la última fase que veremos a continuación.

Etapa 6. **Despliegue:** En esta etapa final se transforma el conocimiento obtenido en acciones concretas dentro del proceso de negocio. Además, se considera esencial documentar y presentar resultados de forma clara y comprensible en un informe con el fin de ser entendido por los usuarios finales.

3. OBJETIVO

Este proyecto busca proporcionar una útil, potente y eficaz herramienta a aquellos usuarios pertenecientes a la industria del turismo, más concretamente, a los gestores de los diferentes alojamientos turísticos con el fin de obtener información acerca de cómo se está gestionando actualmente sus negocios y si se realiza de forma correcta.

Gracias a la aplicación de un modelo algorítmico basado en el análisis de sentimientos a nivel de aspecto, obtendremos un resumen de aquellas características que los clientes de los hoteles valoran de forma positiva y negativa. Esta librería utiliza el modelo de procesamiento natural BERT, una técnica totalmente novedosa y actual que proporciona datos más precisos y mayor contexto debido a la bidireccionalidad en su entreno.

La idea detrás de la elaboración de este trabajo es facilitar la tarea manual del análisis de las reseñas y opiniones que tienen los turistas acerca de un alojamiento en concreto. Por lo que se propone un sistema encadenado que recopile la información proporcionada por cualquier turista de cualquier nacionalidad, es decir, recoger todas las valoraciones independientemente del idioma en el que se encuentren, y generar la valoración de un aspecto puntual de forma totalmente automatizada.

El valor de este proyecto reside en poder simplificar y computarizar todo el proceso de descarga, traducción y análisis de sentimiento a nivel de aspecto de un número ilimitado de reseñas al mismo tiempo, independientemente del origen de los datos o el idioma en el que hayan sido escritos. De esta manera, los gerentes o responsables de cualquier alojamiento turístico en el mundo tendrán a su disposición una gran cantidad de datos e información determinante,

inconcebible con los métodos y técnicas conocidas en la actualidad, para analizar sus estrategias empresariales actuales y tratar de buscar mejoras que conduzcan a un incremento en la satisfacción de sus clientes, rentabilidad y beneficios de la empresa.

4. METODOLOGÍA

En este capítulo se presentan las distintas tareas y decisiones que se han llevado a cabo durante la elaboración del presente trabajo, siguiendo la metodología estándar del proceso que utilizan los expertos en minería de datos descrita en el Marco Teórico, CRISP-DM.

4.1 Comprensión del negocio

En primer lugar, debemos comprender el objetivo del proyecto para ponerlo en consonancia con los del negocio o empresa. En este caso, se trataría de empresas que se dedican al turismo, más concretamente, los alojamientos turísticos. Si tenemos en cuenta las actividades que comprenden esta fase, comenzaríamos por los objetivos comerciales de los hoteles. Para estos, es muy importante la reputación y valor de la marca, pues significaría un aumento clientes potenciales y, por consiguiente, un incremento en las ventas, los ingresos. Uno de los principales actores en la definición de la reputación de un alojamiento turístico son sus usuarios de Internet, aquellas personas que, mediante los medios sociales, expresan libremente su opinión para convencer a otros consumidores de comprar el producto o servicio. En definitiva, el objetivo principal de los hoteles es alojar el máximo número de clientes posibles con la máxima rentabilidad posible; y esto se logra a través de una buena imagen y comentarios positivos.

La segunda de estas tareas comprende la evaluación de la situación actual del negocio. Como ya vimos en capítulos anteriores, las valoraciones de los turistas representan un aspecto muy importante en el turismo, afectando las reservas y compras de futuros clientes. Para el 81% de ellos, las reseñas son el factor más importante a la hora de reservar un hotel (Bhatnagar, 2018).

Actualmente la mayoría de los turistas emplean los medios sociales para expresar su conformidad con el alojamiento tras la estancia, convirtiéndose en una importante fuente de información tanto para consumidores futuros como para los propios alojamientos. Esto significa que los hoteles deben prestar atención a estos comentarios si quieren progresar y crecer en el negocio, pues suponen un impacto directo en sus ingresos. Esto nos lleva a la tercera de las actividades de esta etapa: definir los objetivos a nivel de minería de datos.

El propósito será recopilar toda la información disponible y relevante para los alojamientos turísticos de manera que puedan tomar decisiones en cuanto a la organización y estrategias empresariales. Esto es, no sólo leer les reseñas de los turistas, sino comprenderlas y analizarlas con el fin de aprender más sobre ellos y adecuar los servicios a sus intereses.

Finalmente, entraríamos en la última fase de esta etapa, donde se define el plan del proyecto con sus objetivos, los pasos a seguir y las técnicas a utilizar.

En este trabajo se pretende recopilar las opiniones de los clientes en alojamientos turísticos, para realizar un análisis de sentimientos a nivel de aspecto y obtener aquellas características que valoran positiva y negativamente, con el fin de que esa información sea utilizada para la mejora de la gestión hotelera.

Primeramente, se utilizará la técnica de *web scraping* para extraer la información que nos interesa de la plataforma de viajes TripAdvisor. A continuación, y como las opiniones se encuentran en idiomas diferentes, procederemos a traducirlas al inglés mediante la API de *Google Translate*, para después ejecutar un análisis de sentimientos a nivel de aspecto mediante la librería *aspect-based-sentiment-analysis 2.0.2*. Finalmente, se visualizarán los resultados de cara a los usuarios finales a través de la herramienta de análisis empresarial Power BI.

4.2 Comprensión y preparación de los datos

Seguidamente, debemos detallar cuáles son los datos que vamos a utilizar y cómo vamos a conseguirlos. En este caso, y teniendo en cuenta el objeto de estudio, los valores que nos interesan extraer son las reseñas o comentarios de los diferentes turistas en los alojamientos turísticos de la isla de Gran Canaria, principalmente y, también, de cara a realizar una comparativa, de la isla de Tenerife.

La descripción de estos engloba decisiones como el filtrado de las valoraciones en cuanto a fecha, islas, idioma o número de reseñas. Pero antes que nada debemos definir de dónde los sacaremos y con qué técnica realizaremos la extracción.

Para la extracción de los datos utilizaremos el *web scraping*, una técnica para obtener información de páginas web de forma automatizada. Está compuesto de *bots* que extraen el código HTML de una página y, por consiguiente, los valores almacenados en la base de datos. En nuestro caso, lo utilizaremos para rastrear la plataforma de viajes TripAdvisor y almacenar las valoraciones de los turistas.

Se optó por este sitio web porque es la plataforma turística más grande del mundo, contando con 800 millones de reseñas y comentarios, y 400 millones de visitantes únicos al mes (Conecta Software, s. f.), además de incluir una gran variedad de alojamientos turísticos como son hoteles, apartamentos o casas rurales de toda Gran Canaria.

Asimismo, cada hotel incluía numerosas valoraciones en diferentes idiomas y aportaban mucha información no disponible en otras plataformas, como son el tipo de viaje, la posibilidad de valorar los servicios del hotel de forma separada o escribir un *room tip*.

Antes de comenzar con la extracción debemos considerar qué tipo de datos vamos a descargar y almacenar. Se seleccionaron aquellas reseñas correspondientes al año 2019 por ser período en el que más turistas internacionales visitaron España, con un récord histórico de 83,7 millones de visitantes (Europa Press, 2020). Además, descartamos la idea de escoger aquellas correspondientes al año 2020 ya que el turismo se vio gravemente afectado por la crisis de la COVID-19 y los datos no iban a ser fiables o realistas de cara al objeto de estudio.

Por otro lado, tal y como explicamos al inicio de este epígrafe, se escogieron aquellos alojamientos turísticos pertenecientes a la isla de Gran Canaria y Tenerife, por ser las dos islas más grandes del archipiélago canario y que más viajeros recibe al año, acumulando un total de 4.267.384 y 5.890.704 turistas, respectivamente, en el año estudiado (FRONTUR CANARIAS, 2020).

A continuación, y en cuanto a las reseñas, se seleccionaron aquellos idiomas más comúnmente hablado por los empleados turísticos de las islas, así como las nacionalidades que más visitaron Canarias en ese año. Encabezando la lista se encuentran los británicos, con la cifra de 18 millones de visitantes, seguido de los alemanes con 11,1 millones y los turistas provenientes de Francia (Europa Press, 2020). Al mismo tiempo, el perfil de turista italiano se considera un mercado crucial en las islas, alcanzando la cifra de 455.383 viajeros en 2019 (tourinews, 2020). Por lo tanto, los idiomas escogidos fueron el inglés, alemán, francés, italiano y español, pues las islas reciben mucho turismo nacional también.

Finalmente, tuvimos que preparar los datos para la siguiente etapa, que corresponde con el propio análisis de las reseñas en el que se obtendrán los aspectos positivos o negativos de las opiniones de los turistas. Para ello, se aplicaron técnicas de traducción a través de la API de *Google Translate* para homogeneizarlas a un idioma común, el inglés. Se eligió esta lengua, pues el modelo utiliza técnicas y métodos de análisis que requieren que los datos se encuentren en inglés.

4.3 Modelado

En este apartado se describe el objeto de estudio de este trabajo: el análisis de sentimientos a nivel de aspecto de las valoraciones de los clientes en los alojamientos turísticos de las islas. Todos los datos obtenidos durante las etapas anteriores se incorporan a la técnica analítica para generar resultados que sirvan como una gran fuente de información a los usuarios finales, es decir, los gestores de los hoteles.

Gracias a este análisis y según los resultados obtenidos, los alojamientos podrán realizar acciones concretas dentro del modelo de negocio y establecer conclusiones en términos de cómo se está dirigiendo la compañía y si se está realizando de forma adecuada.

Para ello, se analizarán las reseñas de los turistas en busca de unos aspectos concretos para identificar si hablan positivamente de ellos o negativamente. Esta información será útil para los hoteles para saber qué características valoran más sus clientes y en qué aspectos deben mejorar.

En primer lugar, debemos seleccionar aquellos aspectos que vamos a buscar en las reseñas, deben ser los más relevantes y populares dentro de la industria del turismo. Los hoteles buscan la satisfacción del cliente y esto se consigue ofreciendo unos buenos servicios. Uno de los principales aspectos que reflejan el bienestar de los turistas en su estancia es la limpieza, además de la localización y el servicio ofertado (Stringam et al., 2010). Asimismo, dentro de la categoría de los servicios, se pueden extraer las instalaciones y las vistas (Matilla López et al., 2016).

Por otro lado, un estudio realizado acerca de las opiniones de clientes sobre los hoteles pequeños y medianos de Portugal identifica la habitación, el personal y el desayuno como los aspectos fundamentales y más frecuentemente usados por los turistas en el estudio, con un total del 88,2% (Chaves et al., 2012). Además, dentro del personal, Matilla López (2016) diferencia la recepción como uno de los aspectos para tener en cuenta.

Por consiguiente, los aspectos que hemos tenido en cuenta a la hora de realizar el análisis de sentimientos han sido la recepción, la habitación, el servicio, el personal, la localización, el desayuno, la comodidad, las vistas, la limpieza y las instalaciones.

Estos aspectos, se los pasaremos al algoritmo ABSA, que explicaremos más adelante, para que realice las oportunas predicciones y nos dé como resultado la valoración que tienen los turistas de cada uno de ellos, en los diferentes alojamientos de la isla.

4.4 Evaluación y despliegue

Este apartado corresponde con la última etapa, donde hemos obtenido los resultados finales que debemos proyectar de forma clara y comprensible de cara a los usuarios finales. Estos agentes son los gestores de los alojamientos turísticos, que no buscan saber toda la implementación que hay detrás de los resultados sino la visualización de estos de forma resumida y concisa. Además, descartamos la idea de hacer gráficos y visuales estáticos para aportarle dinamismo a los resultados, y hacer que los gerentes pudieran interactuar con ellos. Para ello, crearemos un panel con la herramienta descrita en el capítulo cuatro: Power BI. Optamos por esta opción pues los paneles son una de las funcionalidades pensadas para la supervisión y obtención de una visión general del negocio. Recopilan las cifras más importantes y relevantes, que pueden provenir tanto de datos locales como de la nube. Además, pueden ser fácilmente compartidos entre los empleados de la organización.

Los gestores de estos alojamientos tienen entonces una potente herramienta con la que podrán transformar los resultados obtenidos en acciones específicas dentro del negocio, incluyendo cambios de estrategias, y mejoras en la gestión o infraestructuras de cara a los futuros visitantes.

5. RECURSOS TECNOLÓGICOS

En este epígrafe se comentarán los diferentes recursos tecnológicos utilizados en el desarrollo del trabajo. Comenzaremos hablando de los recursos software, que incluirán el lenguaje de programación elegido, entorno de desarrollo, paquetes utilizados y algoritmos seleccionados. Para, finalmente, terminar con los recursos hardware, que estarán formados por los equipos informáticos empleados en la ejecución de las aplicaciones anteriores.

5.1 Recursos software

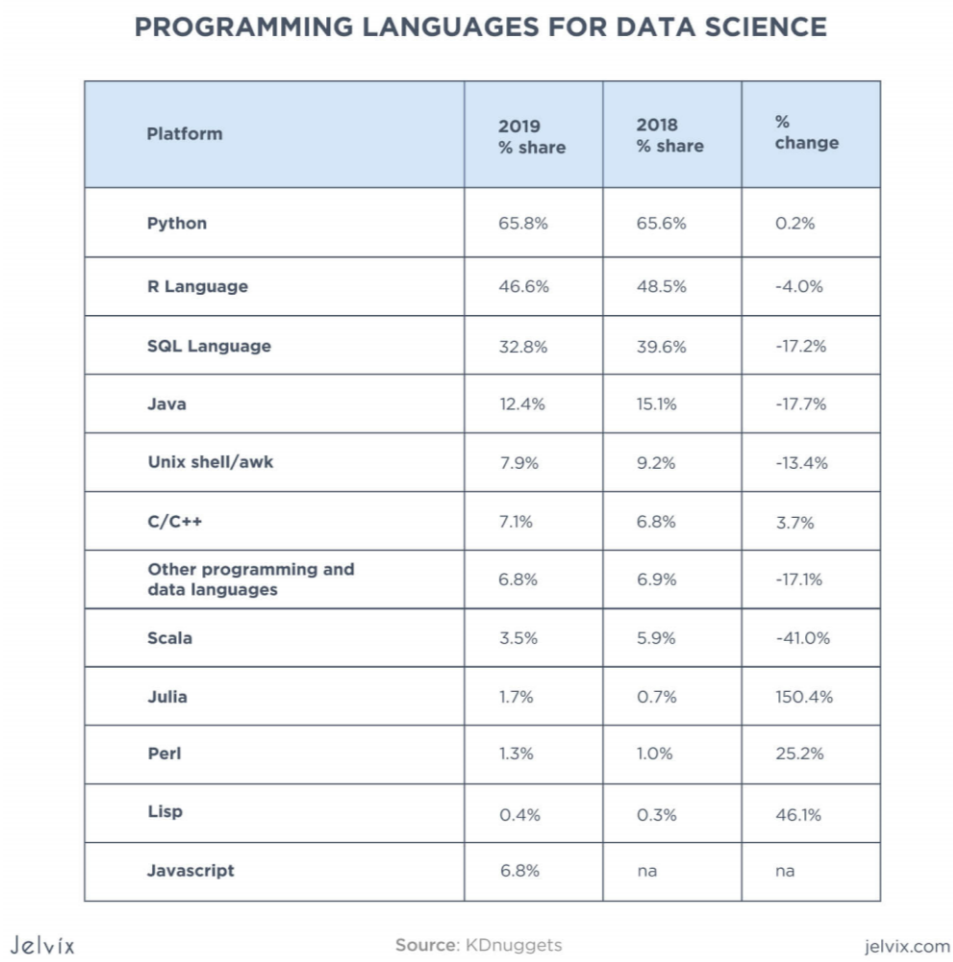
En este apartado se presentarán los diferentes recursos software utilizados para el tratamiento de los datos extraídos de la plataforma de viajes TripAdvisor. Comenzaremos con el lenguaje de programación seleccionado y su justificación, seguido del entorno en el que se ha desarrollado el proyecto, así como los paquetes y APIs utilizados para la obtención de resultados finales.

5.1.1 Elección del lenguaje de programación

Python se considera el lenguaje ideal para iniciarse en la ciencia de datos, pues es rápido, potente y fácil de aprender, además de ser un lenguaje de código abierto, dinámico y con una sintaxis simple. Dispone de más de 137000 librerías esenciales para el aprendizaje automático, la ciencia o manipulación de datos (Andra, 2021). Se trata de un lenguaje universal que permite crear cualquier proyecto, desde aplicaciones sencillas hasta programas de aprendizaje automático (Naumenko, s. f.).

Todas las fuentes consultadas coinciden en que Python es el lenguaje más sencillo e intuitivo, además de ser el más utilizado en todo el mundo. Una evidencia de esto la podemos encontrar en el estudio publicado por KDnuggets, una plataforma online líder en IA, Analytics, Big Data, Data Mining, Data Science, y Machine Learning. Podemos encontrar los datos estadísticos en la tabla 3.

Tabla 3. Top software en analítica y ciencias de datos



PROGRAMMING LANGUAGES FOR DATA SCIENCE

Platform	2019 % share	2018 % share	% change
Python	65.8%	65.6%	0.2%
R Language	46.6%	48.5%	-4.0%
SQL Language	32.8%	39.6%	-17.2%
Java	12.4%	15.1%	-17.7%
Unix shell/awk	7.9%	9.2%	-13.4%
C/C++	7.1%	6.8%	3.7%
Other programming and data languages	6.8%	6.9%	-17.1%
Scala	3.5%	5.9%	-41.0%
Julia	1.7%	0.7%	150.4%
Perl	1.3%	1.0%	25.2%
Lisp	0.4%	0.3%	46.1%
Javascript	6.8%	na	na

Jelvix Source: KDnuggets jelvix.com

Fuente: Naumenko (s.f.)

Además, acorde a una encuesta realizada por JetBrains, compañía de desarrollo software especializada en la creación de herramientas inteligentes, el 85% de desarrolladores de casi 200 países consideraban Python como su principal lenguaje de programación (JetBrains, 2020).

De estos encuestados, el 55% consideraba que el uso principal que le daban a este lenguaje era para el análisis de datos, que coincide con el objeto de este proyecto.

5.1.2 Entorno de desarrollo Google Colab

Google Colab es una herramienta de análisis de datos, basado en el cuaderno interactivo Jupyter, que combina código ejecutable, texto y resultados en un solo documento que se almacena en Google Drive (Tatan, s. f.). Se trata de una herramienta muy cómoda que permite ejecutar programas en Python sin configuración previa y con la posibilidad de compartir el contenido (Google Colaboratory, 2017).

Una de las principales ventajas que presenta este entorno de desarrollo es que el código se ejecuta en la nube, por lo que no hace falta disponer de un ordenador potente, simplemente basta con poseer una cuenta en Google y un navegador web. Para la ejecución de los cuadernos, dispone de tres herramientas hardware cuyas características las podemos encontrar en la tabla 4 (Sharma, 2020).

Tabla 4. Aceleradores por hardware

CPU	GPU	TPU
Procesador Intel Xeon con dos núcleos a 2,30 GHz y 13 GB de RAM	Hasta Tesla K80 con 12 GB de VRAM GDDR5, procesador Intel Xeon con dos núcleos a 2,20 GHz y 13 GB de RAM	TPU en la nube con 180 teraflops de cálculo, procesador Intel Xeon con dos núcleos a 2,30 GHz y 13 GB de RAM

Fuente: Elaboración propia a partir de Sharma (2020)

Otra de las funcionalidades que presenta Google Colab es poder conectar un cuaderno con los documentos, hojas de cálculo o conjunto de datos alojados en tu unidad de Drive.

Con respecto a las librerías de Python, cabe destacar que muchas de ellas como son Pandas, NumPy o Scikit-learn, ya se encuentran previamente instaladas en el entorno.

Una alternativa al uso de Google Colab es Jupyter Notebook, una aplicación de código abierto que permite crear y compartir documentos con código, texto o visualizaciones, pudiendo programar en lenguajes como Python, R, Julia y Scala (Jupyter, 2018). Una de las ventajas que presenta el entorno que hemos utilizado frente a este último, es que no requiere descargar, instalar o ejecutar nada en el equipo local gracias a la utilización de recursos remotos en hosting, mientras que en Jupyter sí que necesitas esos recursos instalados en el equipo.

5.1.3 Paquetes utilizados

En este apartado se detallarán los paquetes y módulos de Python utilizados en el desarrollo del proyecto, así como sus funcionalidades y objetivos. Comprenden librerías empleadas para la extracción de datos además de su estructuración. Los utilizados son:

- Paquete *requests*: su objetivo es hacer que las peticiones HTTP sean más simples y amigables para los humanos, es decir, que la integración con servicios web sea transparente y no exista la necesidad de introducir las consultas manualmente (Reitz, 2013).
- Paquete *bs4*: Beautiful Soup permite extraer datos de archivos HTML y XML, especificando un *parser* responsable de transformar ese documento en un árbol de objetos Python (Richardson, 2020).
- Paquete *pandas*: herramienta de análisis y manipulación de datos de código abierto, flexible y fácil de usar. Ofrece estructuras de datos y operaciones para manipular series temporales y tablas numéricas, pudiendo importar o exportar datos en formatos como csv o xlsx (McKinney, 2008).
- Módulo *time*: proporciona un conjunto de funciones para trabajar con fechas y horas. En este proyecto se ha utilizado la función *sleep()* para interrumpir la ejecución del programa y generar un retardo entre las peticiones (Python Docs, s. f.-c).
- Módulo *random*: genera números pseudoaleatorios. En este caso, se ha utilizado la función *randint()* para devolver un número entero aleatorio dentro de un rango específico (Python Docs, s. f.-b).
- Módulo *json*: convierte objetos Python en una representación serializada JSON y viceversa. Se empleó para transformar las respuestas obtenidas en las peticiones, a un formato más manejable y cómodo (Rico Shmidt, 2013).
- Módulo *math*: conjunto de funciones matemáticas, métodos y constantes (Python Docs, s. f.-a). La función *ceil()* se usó para redondear la división de las valoraciones totales entre el límite establecido por la plataforma (Data, s. f.).

5.1.4 Google Translate API

Una API de traducción sirve para traducir de forma dinámica texto entre dos idiomas. Hoy en día existen muchas de ellas que muchos lenguajes diferentes y que, incluso, pueden detectar el idioma de entrada de un cierto contenido.

Rakuten RapidAPI, un marketplace de APIs para que los desarrolladores las incluyan en sus aplicaciones y proyectos, realizó un estudio comparativo de las diez mejores APIs de traducción del año 2020. Podemos hallar un resumen comparativo de las tres mejores en la tabla 5 que encontramos a continuación (Purkayastha, 2019).

Tabla 5. Mejores APIs para la traducción

API	API Features	Supported Languages	Pricing	Ease of Use
Google Translate API	Translate text, detect the source language	More than 100	Free for 50 requests per day, then $0.05 for each additional request	Easy
Microsoft Translation API	Translate text, detect the source language, transliterate words, bilingual dictionary capabilities	More than 60	Free and varying paid plans from $25 to $200 per month	Easy
Translate API	Translate text, detect the source language	104	Free and varying paid plans from $19 to $59 per month	Easy

Fuente: Purkayastha (2019)

Tras realizar una comparación entre las distintas APIs y observar sus características se ha optado por utilizar la de *Google Translate*. Basada en *Google Cloud*, se considera una de las más utilizadas del mundo, albergando más de 100 idiomas diferentes y pudiendo detectar el idioma de origen cuando este se desconozca.

Cloud Translation permite una traducción rápida y dinámica, con identificación de idiomas automática y alta precisión. Utiliza modelos de aprendizaje automático previamente entrenados, y presenta una integración e implementación sencilla (Google Cloud, s. f.).

Para este proyecto se ha ejecutado la librería gratuita de Python *googletrans*, que implementa la API anteriormente descrita. Se ha escogido esta opción porque puede realizar traducciones masivas, es rápida y fiable, además de gratuita. Utiliza la API Ajax de *Google Translate* para hacer llamadas a métodos como *detect* y *translate* (Han, 2020). Este último será el que más utilicemos, pudiendo detallar los idiomas fuente y de destino. Si no especificamos el lenguaje de origen, la API se encarga de detectarlo automáticamente, así como, si no lo hacemos con el de destino, se traducirá por defecto al inglés.

5.1.5 Sentiment Analysis API

Para la elaboración de este trabajo no nos interesa realizar un análisis de sentimientos simple, puesto que clasificaría la opinión en positiva, negativa o neutra de forma global. Las

conclusiones que queremos sacar de las reseñas de los turistas son los aspectos que valoran positiva y negativamente de los alojamientos.

Si realizáramos un análisis de sentimientos normal utilizaríamos librerías como *TextBlob* o *NLTK-VADER*, pero en este caso tendremos que usar un algoritmo que identifique de forma indirecta unas características que hemos definido previamente.

Para ello utilizaremos el método ABSA o análisis de sentimiento basado en aspectos, cuya definición vimos en el capítulo dos. Es la forma más práctica y real a la hora de analizar y procesar texto en el que se expresan emociones, pues los clientes de los hoteles hablan indistintamente de varias cualidades en una misma valoración. De esta forma se podrá obtener una comprensión más profunda de la satisfacción del cliente con el alojamiento turístico.

La implementación de este tipo de algoritmos se considera todavía un gran reto, pues comprender el lenguaje natural sigue siendo un gran desafío para las máquinas. Actualmente son pocas las APIs que ofrecen un análisis de sentimientos a nivel de aspecto. La mayoría de ellas son de difícil ejecución o de pago.

Sin embargo, hemos descubierto un paquete en Python que resuelve nuestro problema y nos proporciona muchos beneficios a la hora de aplicar sus modelos. El primero de ellos radica en su versatilidad, pues se trata de un servicio adaptable a las necesidades y objetivos de la persona que lo vaya a utilizar. En este caso, lo aplicamos a reseñas de clientes en alojamientos turísticos con los aspectos que estos conllevan, como son el servicio, el personal, el restaurante, las habitaciones… Pero podemos implementarlo en muchos otros casos, como por ejemplo las reseñas de un restaurante o productos, análisis de *tweets*…

Se trata de la librería *aspect-based-sentiment-analysis 2.0.2*, cuya última versión fue lanzada en diciembre del pasado año. Es un paquete de código abierto que incluye modelos de aprendizaje profundo, y proporciona un servicio robusto, estable y escalable.

Como ya sabemos y explicamos anteriormente, para el procesamiento de lenguaje natural o NLP, necesitamos un *pipeline*. Este se devuelve en la primera función que hay que ejecutar (ecuación 2), junto con la descarga de un modelo en el directorio del paquete y la configuración de TensorFlow, plataforma de código abierto para el aprendizaje automático.

Ecuación 2. Función de configuración del pipeline

$$nlp = absa.load()$$

Fuente: Elaboración propia

A esta función podemos pasarle el modelo que queremos usar explícitamente o utilizar el que viene por defecto (modelo BERT). En nuestro caso hemos escogido esta segunda opción por el potencial que tiene, tal y como explicamos en el epígrafe dos. Este clasificador de sentimientos, como hemos mencionado anteriormente, se basa en la arquitectura del *Transformer*, donde las capas de autoatención contienen la mayoría de los parámetros.

Además de poder incluir el modelo que queramos, nos da la posibilidad de dividir el texto en fragmentos más pequeños con el fin de procesar las oraciones de manera independiente. Esto se realiza pasándole a la función la variable *text_splitter*, que comprende el modelo spaCy, librería de software para procesamiento de lenguajes naturales.

Los detalles de cómo se implementa y ejecuta el *pipeline,* así como las predicciones que puede hacer el modelo los veremos con más detenimiento en el capítulo del desarrollo (epígrafe 6).

En definitiva, el análisis de sentimientos basado en aspectos puede ser muy ventajoso para diferentes negocios, pues nos ayuda a comprender los gustos de los clientes hacia una marca, analizar los productos y servicios para cambiar las estrategias empresariales o compararlos con los de la competencia, o seguir la evolución de los consumidores ante ciertas características o atributos de un producto/servicio.

5.1.6 Power BI

Power BI es un servicio de análisis empresarial, dado a conocer por la compañía Microsoft en septiembre de 2013. Sin embargo, no fue hasta dos años más tarde cuando se lanzó al público general. Se trata de un sistema basado en la nube, por lo que permite el acceso a los datos tanto fuera como dentro de la organización y desde cualquier dispositivo.

Es conocida como una de las herramientas más novedosas y potentes en el mundo de la Inteligencia Empresarial, muy fácil e intuitiva en su utilización (Quonext, s. f.). Su objetivo reside en proporcionar visualizaciones interactivas con una interfaz lo suficientemente simple como para que los usuarios finales la entiendan y puedan crear sus propios informes.

Power BI es un software como servicio (SaaS) que ofrece cuadros de mando interactivos, creados y actualizados a partir de muchas fuentes de datos diferentes (Negrut, 2018). Estos cuadros de mando proporcionan información importante y relevante de los datos de la empresa mediante la creación de informes con gráficos descriptivos.

Una de las principales ventajas que presenta esta tecnología es la interactividad entre los usuarios y distintos trabajadores de la empresa, pues permite compartir conocimientos y colaborar en informes en tiempo real, para tomar decisiones conjuntas que lleven a una mejora de las estrategias empresariales. El proceso de la toma de estas decisiones puede ser llevado a cabo de mejor manera y más consensuada gracias a la visualización de los datos. El efecto visual que tienen los informes y las gráficas que aparecen en ellos ayuda a interpretar datos complejos.

Power BI consiste en una solución analítica de negocios que se centra en tres áreas principales: preparación y análisis de los datos, visualizaciones, y colaboración e intercambio.

La primera de estas tareas se centra en conectar el software con los datos o archivos de origen y prepararlos para tener datos correctamente estructurados y pasar a su posterior análisis, en el que se realizarán cálculos sobre esos datos.

Posteriormente encontramos las visualizaciones, donde se crean los distintos paneles e informes con sus respectivos gráficos e imágenes. No sólo se disponen de visuales predeterminados, sino que los usuarios tienen la posibilidad de crear el suyo propio. Además, dentro de los paneles, los gráficos están conectados entre sí, promoviendo la interacción entre datos.

Por último, hablamos de la colaboración e intercambio mencionado anteriormente. La posibilidad de trabajar en equipo y compartir los resultados e informes tanto dentro como fuera de la compañía y en tiempo real.

Power BI no es una herramienta única, sino que engloba un conjunto de herramientas múltiples que interactúan entre sí. Los tres componentes principales son los siguientes:

- Power BI Desktop es una aplicación local o de escritorio que se puede descargar gratuitamente y está pensada para usuarios que requieran una solución de *Business Intelligence* sin ser expertos en la materia. De las tres áreas que vimos previamente, incluye la transformación de datos, la visualización de estos, y la creación de informes.
- Power BI Service es el servicio en línea (SaaS) con una funcionalidad parecida al anterior. Se diferencian en que este sistema permite publicar y compartir informes con el resto de los empleados de la organización, además de poder configurar la actualización de datos automáticamente para que los usuarios siempre dispongan de la última información.

- Power BI Mobile es la aplicación móvil para visualizar los paneles e informes desde cualquier parte y en cualquier momento. Actualmente se encuentra disponible para Windows, iOS y Android.

La naturaleza de los datos con los que podemos conectar esta herramienta es muy diversa. El origen de estos puede provenir de la nube o del entorno local, e incluyen Dynamics 365, Azure SQL Database, Salesforce, Excel, SharePoint, entre otros (Microsoft Docs, 2021). Además, estos datos se pueden cruzar entre sí, de manera que en un mismo panel o informe podemos tener información proveniente de sitios distintos.

Por otro lado, es interesante nombrar una de las funcionalidades más atractivas de este recurso tecnológico como es la incorporación de sistemas inteligentes. Gracias a la inteligencia artificial, los usuarios podrán predecir tendencias y resultados futuros, así como aplicar algoritmos de aprendizaje profundo en la preparación de los datos. Algunos de estos servicios incluyen el análisis de sentimiento, extracción de frases clave, detección de idioma y etiquetado de imágenes.

Power BI ha demostrado ser líder en su campo después de ser considerada, durante catorce años consecutivos, la mejor plataforma para el análisis de datos y la inteligencia empresarial en el Cuadrante Mágico de Gartner, una herramienta que examina la innovación y desarrollo de empresas tecnológicas a nivel mundial, tal y como observamos en la ilustración 16.

Es indiscutible la cantidad de beneficios y aportaciones que ha hecho Power BI a los negocios y, en general, al análisis de datos. Su capacidad de adaptación y tendencia visionaria ante los constantes cambios en esta área han hecho a Power BI merecedora de este reconocimiento.

Las empresas que utilicen estas herramientas en su negocio podrán disponer de una gran ventaja competitiva frente al resto de compañías de su sector, pues gracias a la previsión de resultados, podrán detectar nuevas oportunidades, modificando sus estrategias empresariales. Además, se trata de un recurso altamente versátil que puede ser utilizado en los diferentes departamentos de la compañía, pudiendo adaptar los cuadros de mando, y permitiendo el intercambio de información entre ellos.

Ilustración 16. Cuadrante mágico de plataformas de análisis e Inteligencia Empresarial

Fuente: Gartner (2021)

5.2 Recursos hardware

5.2.1 Ordenador personal

Todos los desarrollos, tanto para la descarga de datos mediante *web scraping*, como para las conexiones con las distintas APIs de traducción y análisis de sentimientos, han sido realizados en Python en un ordenador personal con las características descritas en la tabla 6.

Tabla 6. Especificaciones ordenador personal

Características	
Sistema Operativo	Windows 10 Home 64 bits
Procesador	Intel® Core™ i7-1165G7
Memoria RAM	16,0 GB
Disco Duro	512 GB
Tarjeta Gráfica	Intel® Iris Xe Graphics; NVIDIA GeForce MX350

Fuente: Elaboración propia

6. DESARROLLO

Durante este capítulo se detallarán las tareas y procesos realizados para la obtención de datos, así como su tratamiento y posterior análisis. Se comentarán los objetivos, módulos y librerías utilizadas en cada una de ellas, además de los algoritmos llevados a cabo.

6.1 Extracción de los datos

En este apartado se explica cómo se obtienen los datos tanto de las reseñas de los turistas de cada hotel como de la información correspondiente a cada alojamiento, mediante el uso de las librerías explicadas en el capítulo 5.

6.1.1 Enlaces de los alojamientos

Para la extracción de la información principal de este trabajo, es decir, las valoraciones de los clientes, se ha empleado el mencionado *web scraping*, una técnica para obtener datos de sitios web mediante programas software.

En primer lugar, y a través de la plataforma de viajes TripAdvisor, se buscaron los alojamientos disponibles en la isla de Gran Canaria. Había que tener en cuenta el número de páginas de alojamientos para poder rastrearlas todas y conseguir la URL de cada hotel, tal y como observamos en la ilustración 17. Al final obtendríamos una lista con todos los enlaces para, posteriormente, entrar en cada uno de los alojamientos y conseguir sus reseñas. Esta lista la hemos llamado *hotels_links*.

Ilustración 17. Extracción de los enlaces de los hoteles

```
[ ]  # Comprobamos cuántas páginas hay para rastrearlas todas
     hotel_pages = 46
     # Función para, según la página en la que estemos, coger una url u otra
     def get_hotel_url(page):
       if page == 0:
         return 'https://www.tripadvisor.com/Hotels-g187471-Gran_Canaria_Canary_Islands-Hotels.html'
       return f'https://www.tripadvisor.com/Hotels-g187471-oa{page*30}-Gran_Canaria_Canary_Islands-Hotels.html'
```

```
[ ]  hotels_links = []
     for i in range(hotel_pages):
       # Retardo aleatorio para evitar bloqueos
       time.sleep(random.randint(2, 8))
       # Descargamos html
       hotels_html = requests.get(get_hotel_url(page=i))
       # Comprobamos que la respuesta sea válida
       assert hotels_html.status_code == 200
       # Analizamos el html
       hotels_soup = BeautifulSoup(hotels_html.text, 'html.parser')
       # Cogemos los enlaces de cada hotel
       links = hotels_soup.select('.listing')
       for link in links:
         hotels_links.append('https://www.tripadvisor.com'+link.a['href'])

     hotels_links
```

Fuente: Elaboración propia

En la parte inicial del código observamos que según cambiamos de página, la URL se modifica, añadiendo "oa*X*" a la dirección web original. Esa *X* representa el número de la página en la que estamos, multiplicada por 30. De esta forma, si por ejemplo seleccionamos la número 4, obtendríamos el siguiente enlace: https://www.tripadvisor.com/Hotels-g187471-oa90-Gran_Canaria_Canary_Islands-Hotels.html, teniendo en cuenta que el índice empieza en 0. Por lo que, en este caso, estaríamos multiplicando 3*30.

Así es como lograríamos tener los enlaces de cada página, pero nos faltaría obtener la URL de cada alojamiento. El primer paso para ello es mediante una solicitud o petición a través de la librería *requests*, donde se obtiene un objeto *Response* con el contenido de la página web.

Con el fin de no colapsar estas peticiones se ha añadido un retardo aleatorio de entre 2 y 8 segundos, además de una comprobación del código de respuesta. Nos aseguramos de que esta tenga el valor 200, que representa una solicitud correcta e implica la devolución de la página solicitada de forma exitosa.

Por otro lado, se hace uso de la librería Beautiful Soup para analizar el contenido HTML de la página y transformarlo en un árbol complejo de objetos Python.

A continuación, y gracias a la función *select()* de la librería podemos buscar etiquetas de CSS y guardar su contenido en una variable. En este caso, intentábamos localizar la correspondiente a la clase *listing* (en Beautiful Soup, para expresar que se trata de una etiqueta de clase añadimos un punto al principio, "*.listing*"). Esta etiqueta se encontraba en los nombres de los alojamientos que encontramos en cada página, que, a su vez, se trataban de hipervínculos que nos guiaban al enlace individual de cada uno de los hoteles.

No obstante, para poder acceder a estos de forma correcta, tuvimos que añadir la primera parte correspondiente a la URL de la plataforma de viajes: https://www.tripadvisor.com/.

Una de las principales ventajas que presenta la librería Beautiful Soup es poder acceder a cualquier elemento de manera sencilla mediante el uso de signos como los puntos o corchetes, como en este caso.

Tras este paso y tener todos los enlaces de los alojamientos en una lista, recorreremos cada uno de ellos en busca de nuestro objetivo principal: las reseñas de los turistas.

6.1.2 Valoraciones de los turistas

Al haber obtenido ya cada uno de los enlaces de los alojamientos, podemos proceder a la extracción de las valoraciones de los turistas con las características descritas en el apartado tres.

Para disponer de valores más relevantes y que reflejen autenticidad en las opiniones que tienen los viajeros acerca de un alojamiento, se han descartado aquellos hoteles con menos de cinco comentarios, mediante la introducción de una variable *count*.

Al contrario que para los enlaces de hoteles, para esta ejecución se utilizó GraphQL, un lenguaje de consulta y manipulación de datos. Se manejó esta técnica ya que se podían obtener muchos más datos interesantes y significativos para el estudio, no visibles a simple vista en la página web.

Además, tras guardar la consulta en una variable *response*, se podía acceder de manera muy intuitiva y fácil a cualquier dato que nos interesara y de cualquier nivel simplemente escribiendo la variable entre corchetes, tal y como vemos en la siguiente ilustración.

Ilustración 18. Ejemplo de uso GraphQL

```
reviews = response['reviewListPage']['reviews']
```

Fuente: Elaboración propia

Esto es simplemente un ejemplo de cómo acceder y almacenar todas las reseñas de un alojamiento turístico en un variable gracias al uso del lenguaje de consulta GraphQL, que nos permite obtener muchos recursos en una sola solicitud.

Ahora que tenemos todas las valoraciones en una única variable, recorreremos la lista para ir guardando los datos que nos interesen de cada una de ellas. Como paso previo, tal y como comentamos anteriormente, utilizaremos la fecha de creación de la reseña para filtrar por aquellos comentarios que se hicieron durante el año 2019 y descartar el resto, tal y como observamos en la ilustración 19.

A través de la respuesta obtenida en la petición, pudimos acceder a los diferentes componentes de las opiniones. Todas estas no tenían siempre las mismas características, es por ello por lo que siempre teníamos que comprobar que existiera el elemento, y de no ser así, lo dejaríamos en blanco.

Ilustración 19. Extracción de los datos de las valoraciones

```
or review in reviews:
 date = review['createdDate'].split('-')
 year = int(date[0])
 if year != 2019:
   continue
 count = count+1
 review_title = review['title']
 review_description = review['text']
 island = review['location']['parent']['additionalNames']['normal']
 review_data = {
   'Hotel Name': hotel_name,
   'Review Date': review['createdDate'],
   'Stay Date': review['tripInfo']['stayDate'] if review['tripInfo'] is not None else None,
   'Island': island,
   'Lang': review['language'],
   'Room Tip': review['roomTip'] if 'roomTip' in review else None,
   'Review Title': review_title,
   'Review': review_description,
   'Review Stars': review['rating'],
   'Trip type': review['tripInfo']['tripType'] if review['tripInfo'] is not None and review['tripInfo']['tripType'] is not None else None
   'User Name': review['userProfile']['displayName'] if review['userProfile'] is not None else None,
   'Hometown': review['userProfile']['hometown']['location']['additionalNames']['long'] if review['userProfile'] is not None and review['
   'Hotel Response': review['mgmtResponse']['text'] if review['mgmtResponse'] else None,
   'Response User': review['mgmtResponse']['username'] if review['mgmtResponse'] else None,
   'Response Date': review['mgmtResponse']['publishedDate'] if review['mgmtResponse'] else None
```

Fuente: Elaboración propia

Toda esta información se guardó en una estructura de datos DataFrame de la librería Pandas, y se almacenó en un CSV, para poder representar los datos en formato tabla de manera que quedara más visual y organizado.

Tabla 7. Resultados de las valoraciones

	Hotel Name	Review Date	Stay Date	Island	Lang	Room Tip	Review Title	Review	Review Stars	Trip type	User Name	Hometown	Hotel Response	Response User	Response Date	Value Stars	Rooms Stars	Location Stars	Clear
0	Hotel Faro, a Lopesan Collection Hotel	2019-12-19	2019-12-31	Gran Canaria	en	None	Still working	Thought people might be interested in learning...	4	COUPLES	Arthur B	London, United Kingdom	None	None	None	NaN	NaN	NaN	
1	Hotel Faro, a Lopesan Collection Hotel	2019-07-15	2019-07-31	Gran Canaria	en	None	Building site	I have just returned from Gran Canaria and I c...	1	COUPLES	DPL	None	None	None	None	NaN	NaN	NaN	
2	Hotel Faro, a Lopesan Collection Hotel	2019-06-03	2019-05-31	Gran Canaria	en	None	Fanstastic stay	The hotel is located in one of the best place ...	5	FRIENDS	sssixt	None	Dear Sssixt,\nFirstly we would like to thank y...	MontseF303	2019-06-04	5.0	5.0	5.0	

Fuente: Elaboración propia

La tabla 7 muestra los resultados que se obtendrían tras la ejecución del código *web scraping*. Tal y como explicamos anteriormente, algunas reseñas no contemplaban toda la información todos los aspectos que queríamos guardar, es por ello por lo que podemos apreciar que esos datos toman valores como *None* o *NaN*.

Toda esta implementación se realizó en el cuaderno *TripAdvisor_Scraping(GC).ipynb* y sus resultados se almacenaron en el archivo *reviews_2019(GC).xlsx*. Optamos por utilizar un Libro Excel para guardar los resultados para almacenar texto enriquecido y darles a los datos un formato de tabla más visual.

6.1.3 Datos de los hoteles

Aparte de las valoraciones de los clientes de los hoteles, consideramos necesario incluir un documento con todas las características de cada uno de los alojamientos turísticos. Para ello, se implementó otro cuaderno Jupyter (*Hotel_Data(GC).ipynb*) en el que recolectamos datos tanto con Beautiful Soup, como con GraphQL. A continuación, en la ilustración 20, encontramos un ejemplo de cada uno:

Ilustración 20. Ejemplo de Beautiful Soup y GraphQL

```
hotel_data = []
for hotel_url in hotels_links:
  # Retardo aleatorio para evitar bloqueos
  time.sleep(random.randint(2, 8))
  # Descargamos el html del hotel
  hotel_html = requests.get(hotel_url)
  # Comprobamos que la respuesta sea válida
  if hotel_html.status_code != 200:
    continue
  # Analizamos el html
  hotel_soup = BeautifulSoup(hotel_html.text, 'html.parser')

  #GraphQL
  response = request_graphql(hotel_url)[1]['data']['locations'][0]

  # Cogemos el nombre del hotel
  hotel_name = response['name']
  print(hotel_name)

  # Estrellas
  if hotel_soup.find("span", class_="cwu1UFvH"):
    hotel_stars = hotel_soup.select('.cwu1UFvH')[0].svg['title']
    hotel_stars = hotel_stars.split(" ")[0]
  else:
    hotel_stars = None
  print(hotel_stars)
```

Fuente: Elaboración propia

Asimismo, y con el fin de recolectar más información, se añadieron funciones para obtener el script de la página web, el código fuente (ilustración 21). Esto se incluyó para recoger datos no accesibles a través de los métodos utilizados hasta ahora. Algunos de estos valores comprendían: las características de las habitaciones, servicios de los alojamientos, su descripción y los idiomas hablados por el personal del hotel.

Ilustración 21. Funciones para obtener script de la página

```
def get_data(soup):
  all_scripts = soup.find_all('script')
  # Buscamos el script que tenga 'WEB_CONTEXT' dentro, que es el JSON todos los datos
  script = list(filter(lambda s: 'WEB_CONTEXT' in s.text, all_scripts))[0]
  script = script.text.replace('window.__WEB_CONTEXT__={pageManifest', '{"pageManifest"')
  end_of_json = ';(this.$WP'
  assert(end_of_json in script)
  index = script.index(end_of_json)
  script = script[:index]
  datos = json.loads(script)['pageManifest']['urqlCache']
  return datos
```

```
def get_amenities(soup):
  datos = get_data(soup)
  amenities = get_recursively(datos, 'hotelAmenities')[0]
  property_amenities = amenities['highlightedAmenities']['propertyAmenities'] + amenities['nonHighlightedAmenities']['propertyAmenities']
  return property_amenities
```

```
def get_features(soup):
  datos = get_data(soup)
  amenities = get_recursively(datos, 'hotelAmenities')[0]
  room_features = amenities['highlightedAmenities']['roomFeatures'] + amenities['nonHighlightedAmenities']['roomFeatures']
  return room_features
```

Fuente: Elaboración propia

En este paso debemos tener en cuenta tres funciones. La primera de ellas (*get_data()*) se utiliza para almacenar el script, analizarlo y convertirlo en un diccionario Python gracias al método *loads()*. Seguidamente, nos encontramos *get_recursively()*, una función recursiva predefinida para encontrar un valor en un diccionario anidado. En este caso, lo que estamos haciendo es pasarle los datos del script y el valor que queremos buscar. Por último, encontramos las propias funciones de los elementos que queremos buscar y guardar, *get_amenities()* y *get_features()* son sólo algunos ejemplos. En ellas tenemos en cuenta tanto las facilidades que se muestran directamente (*highlighted*), como las que se encuentran ocultas (*nonHighlighted*) y debemos presionar "Mostrar más" para visualizarlas.

Toda la información anterior se almacenó en una nueva estructura de datos DataFrame, que podemos visualizar en la tabla 8. Además, también hemos seguido el mismo criterio que en las valoraciones y hemos descartado la información de los hoteles que tienen menos de cinco reseñas.

Tabla 8. Características de los alojamientos

	Hotel Name	Hotel Class	Description	Location	Phone	Price	Amenities	Room Features	Number of Rooms	Number of Reviews	General punctuation	Spoken Languages
0	Hotel Faro, a Lopesan Collection Hotel	4.0	The name of this five stars hotel is taken fro...	Plaza de Colon 1, 35100, Maspalomas, Gran Cana...	None	$249	[Paid public parking nearby, Wifi, Hot tub, Fi...	[Blackout curtains, Bathrobes, Air conditionin...	179	1007	4.5	[English, French, Spanish, German]
1	AxelBeach Maspalomas	3.0	The new AxelBeach Maspalomas - Apartments & Lo...	Avenida Tirajana 32 Entrada Por C/ Timple, 351...	None	$104	[Free High Speed Internet (WiFi), Pool, Fitnes...	[Air conditioning, Housekeeping, Private balco...	92	2542	4.5	[English, French, Spanish, German, Italian]
2	Santa Catalina, a Royal Hideaway Hotel	5.0	Surrounded by the island's hundred-year-old ve...	Calle del Leon y Castillo 227 Ciudad Jardín, 3...	1 (833) 422-3631	$136	[Electric vehicle charging station, Free High ...	[Bathrobes, Air conditioning, Housekeeping, In...	204	284	4.5	[English, French, Spanish, German, Italian]

Fuente: Elaboración propia

Esta implementación se realizó en el cuaderno *Hotel_Data(GC).ipynb* y sus resultados se almacenaron en el archivo *hotels_info(GC).xlsx*.

6.2 Preparación de los datos

Antes de realizar el análisis de sentimientos a nivel de aspecto debemos asegurarnos una uniformidad en los datos, de manera que todos ellos puedan ser reconocidos por el algoritmo. Esto es debido a que la librería utiliza un paquete *pipeline* preentrenado en inglés.

Por lo tanto, todas nuestras reseñas deberán estar en un idioma común (el inglés), de manera que se obtengan los datos finales con los que se trabajarán y aplicarán los posteriores modelos para así garantizar el correcto funcionamiento del análisis de aspectos.

6.2.1 Traducción de las valoraciones

Para la traducción de las opiniones de los turistas hemos optado por utilizar la librería *googletrans* pues, como comentamos en el apartado de Recursos Tecnológicos, es una herramienta gratuita, fácil de utilizar y fiable, además de rápida.

Para poder empezar a utilizar la API de *Google Translate* debemos instalarla haciendo uso del comando *pip*, utilizado para la gestión e instalación de paquetes software escritos en Python (ilustración 22). En este caso, hemos instalado la última versión porque, de lo contrario, nos daría un error y no podríamos hacer uso de la API.

Ilustración 22. Instalación de Google Translate API

```
!pip install googletrans==4.0.0-rc1
```

Fuente: Elaboración propia

A continuación, lo primero que debemos hacer es inicializar el traductor en sí. Para ello, lo importaremos desde el módulo *googletrans*, y creamos un objeto de la clase *Translator*.

Ahora simplemente podremos hacer uso de la herramienta mediante la llamada al método *translate()*, que recibe como parámetro el texto que queremos traducir. Esta función también nos permite especificar los idiomas de origen y destino como parámetros, mediante el uso de las variables *src* y *dest*, respectivamente. Si, en cambio, no declaramos idioma de entrada, la librería lo detectará de forma automática. Al igual que si no especificamos idioma de salida, la traducción se realizará al inglés por defecto.

Además, con este método podremos obtener no sólo la traducción del texto sino otros atributos como son los idiomas de origen y destino, el texto original o la pronunciación.

Ilustración 23. Traducción de las valoraciones

```
#Recorremos las filas del archivo
for i, row in df.iterrows():
  # Retardo aleatorio para evitar bloqueos
  time.sleep(2)
  # Sólo traducimos los que no están en inglés
  if row['Lang'] == 'en': continue
  print('Lang:', row['Lang'])

  print('Antes: ', row['Room Tip'])
  if row['Room Tip'] is None:
    trans_tip = None
  else:
    try:
      trans_tip = translator.translate(row['Room Tip'], dest='en').text
    except:
      trans_tip = None
  print('Después:', trans_tip)
  df['Room Tip'] = df['Room Tip'].replace(row['Room Tip'], trans_tip)

  print('Antes: ', row['Review Title'])
  try:
    trans_title = translator.translate(row['Review Title'], dest='en').text
  except:
    trans_title = None
  print('Después:', trans_title)
  df['Review Title'] = df['Review Title'].replace(row['Review Title'], trans_title)
```

Fuente: Elaboración propia

Para nuestro proyecto, comenzamos recorriendo las filas del archivo *reviews_2019(GC).xlsx* que contienen todas las reseñas de Gran Canaria en el año 2019. Como vamos a traducirlas al inglés, aquellas que ya se encuentran en ese idioma, las ignoramos mediante el uso del comando *continue*, tal y como observamos en la ilustración 23. Después, traducimos la columna que nos interese y sustituimos el nuevo valor en el archivo original a través de la función *replace()* de la librería Pandas. Esto lo haremos utilizando la estructura *try/except* para asegurarnos de que no se generen errores o excepciones a la hora de traducir y quedarnos con aquellos datos aptos y relevantes, garantizando la limpieza de valores para la siguiente fase de análisis.

Las columnas traducidas corresponden con los consejos sobre las habitaciones (*Room Tip*), el título de la valoración (*Review Title*), la propia opinión del turista (*Review*), y la respuesta del alojamiento (*Hotel Response*). Se han escogido estos valores pues el resto de los datos no precisan traducción, como son el nombre del hotel, del usuario, la localización, entre otros.

A continuación, encontramos un ejemplo de traducción cuyo idioma de origen es el francés. Tal y como apreciamos en la imagen, realiza correctamente las traducciones de los tres primeros atributos, que son los consejos sobre la habitación, el título de la reseña, y la

valoración en sí; pero no traduce la respuesta del alojamiento porque, en este caso, no se identificaba ninguna.

Ilustración 24. Ejemplo de traducción

```
Lang: fr
Antes:  Un peux excentré du Yumbo mais globalement intéressant pour glander autour d'une piscine
Después: An eccentric of yumbo but overall interesting for glander around a swimming pool
Antes:  Bon établissement réservé aux gays; doit prévoir des travaux d'amélioration
Después: Good school reserved for gays;must provide improvement work
Antes:  Nous passons 15 jours une à deux fois par ans dans cet établissement. Bon niveau de prestat
Después: We spend 15 days one to twice a year in this establishment.Good level of service for a pri
Antes:  In
Después: In
```

Fuente: Elaboración propia

Finalmente, almacenaremos estos resultados en el archivo original, tal y como explicamos previamente, pero le cambiaremos el nombre a *translated_reviews_2019(GC).xlsx* para diferenciarlo del Excel de las valoraciones.

6.3 Análisis de sentimientos a nivel de aspecto

Tal y como explicamos en apartados anteriores, estudiaremos las valoraciones de los clientes de alojamientos turísticos en el nivel de aspecto que es el más profundo del análisis de sentimientos. Esto lo haremos utilizando la librería *aspect-based-sentiment-analysis*, en su última versión, la 2.0.2.

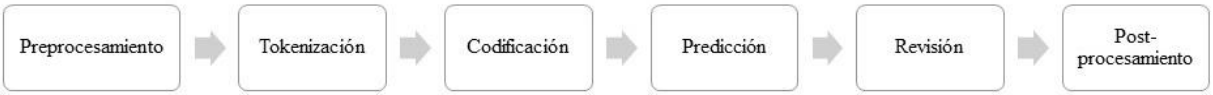
Esta librería, engloba un modelo *Transformer*, concretamente el modelo BERT; además, utiliza la librería TensorFlow para el aprendizaje automático y SpaCy para el procesamiento de lenguaje natural (Rolczynski, 2020). En definitiva, se basa en la *pipeline* de NLP, descrita en el capítulo dos de Marco Teórico.

De manera simple se podría decir que el algoritmo está formado por el modelo lingüístico BERT, que proporciona las características o aspectos, y un clasificador lineal.

En línea con lo mencionado, debemos entender cómo funciona este algoritmo y de qué manera hace uso del *pipeline*, preparando las entradas y sabiendo interpretar las salidas. Procesa enormes conjuntos de datos para averiguar las relaciones entre las palabras. Dentro del procesamiento de las entradas, debemos convertir del texto mediante el proceso de tokenización y codificación para después pasárselo al modelo. Como resultado, es capaz de codificar palabras, cadenas sin sentido, en vectores enriquecidos con información.

Una de las grandes diferencias que presenta este algoritmo frente a otros es que incluye una revisión tras la predicción, tal y como apreciamos en la siguiente figura:

Ilustración 25. Etapas del pipeline



Fuente: Elaboración propia a partir de Rolczyński (2020)

Durante esta fase, un componente independiente llamado *professor*, supervisa y explica la predicción de un modelo. Este elemento se ha introducido para revisar los estados internos del modelo y proporcionar explicaciones de las predicciones que, esporádicamente, revelan comportamientos inesperados (Rolczyński, s. f.). Se compone de modelos auxiliares que examinan el razonamiento del modelo y corrigen cualquier debilidad del mismo, obteniendo un mayor control sobre el proceso de toma de decisiones.

En este caso, se hace uso de un clasificador auxiliar que predice si un texto está relacionado con un aspecto o no. Si no existe referencia a un aspecto en un texto, la variable *professor* establece un sentimiento neutro, independientemente de la predicción del modelo. Este clasificador es el *Basic Reference Recognizer*, que comprueba si un aspecto está claramente mencionado en un texto. Representa un texto y un aspecto como dos vectores, mide la similitud del coseno entre ellos y, posteriormente, utiliza la regresión logística simple para hacer una predicción. Calcula las predicciones del texto y los aspectos, sumando los vectores de *subtokens* y los *embeddings* independientes del contexto, provenientes de la capa de *embedding*, vista en el epígrafe dos.

A continuación, en la ilustración 26, podemos observar una descomposición de las etapas del *pipeline*, con sus diferentes métodos y parámetros. Vemos que, en cada paso, se recogen los resultados de la fase anterior para producir las siguientes salidas.

Ilustración 26. Descomposición del pipeline

```
task = nlp.preprocess(text=..., aspects=...)
tokenized_examples = nlp.tokenize(task.examples)
input_batch = nlp.encode(tokenized_examples)
output_batch = nlp.predict(input_batch)
predictions = nlp.review(tokenized_examples, output_batch)
completed_task = nlp.postprocess(task, predictions)
```

Fuente: Rolczyński (2020)

Para empezar a utilizar cualquier librería preexistente, lo primero que debemos hacer es instalarla mediante el comando *pip* y, en este caso, asegurarnos de que utilizamos la última versión especificando el número mediante "==".

Como hemos comentado, esta librería se fundamenta en la *pipeline* de NLP, por lo que primeramente y antes de realizar cualquier otra acción, debemos llamar a la función que configura el *pipeline* listo para usarse. A este método se le puede pasar el modelo que vayamos a utilizar en forma de parámetro. Si no lo hacemos, cogerá el que proporciona el paquete por defecto, que en este caso es el modelo BERT. Utilizamos este modelo porque podemos beneficiarnos directamente de la predicción de la siguiente frase para formular la tarea como una clasificación de pares de secuencias (*text subtokens* y *aspect subtokens*).

Otros de los parámetros aceptados es el *text_splitter*, que almacena el resultado de aplicar la función *sentencizer()*. En muchas ocasiones, el texto es tan largo que el sentimiento que recoge tiende a ser confuso. Por ello, se recomienda dividirlo en fragmentos más pequeños mediante el uso del modelo SpaCy CNN, que separa un documento en oraciones más simples de manera que puedan ser procesadas de forma independiente. En este caso sí que lo hemos pasado como un parámetro de la función (ilustración 27) pues la mayoría de las valoraciones englobaban un conjunto de frases que debíamos separar.

Ilustración 27. Inicialización del pipeline

```
df = pd.read_excel('/content/drive/MyDrive/TFG/Resultados/translated_reviews_2019(GC).xlsx')

#Dividir texto en fragmentos más pequeños
sentencier = absa.sentencizer()
#Inicialización del pipeline
nlp = absa.load(text_splitter=sentencier)
#Aspectos a buscar
aspects = ["reception", "room", "service", "staff", "location", "breakfast", "comfort",
           "view", "cleanliness", "facilities"]
df.head()
```

Fuente: Elaboración propia

En primer lugar, leemos el archivo que contiene las reseñas que queremos analizar, seguido de la inicialización del separador de texto, que pasamos como parámetro a la función *load()* para comenzar a utilizar el modelo. Además, se crea una lista con aquellos aspectos que queramos localizar en las opiniones de los turistas. En este caso, nos hemos basado en las investigaciones y estudios mencionados anteriormente para seleccionarlos.

Ilustración 28. ABSA

```
hotel_absa = []
#Añadimos barra de progreso al bucle
for i, row in tqdm(df.iterrows(), total=df.shape[0]):
  #Almacenamos la reseña
    text = (row['Review'])
    try:
      #Pasamos el texto y aspectos al modelo
      reception, room, service, staff, location, breakfast, comfort, view,
      cleanliness, facilities = nlp(text, aspects)
    except:
      continue
    #Añadimos los resultados
    hotel_absa.append({
      'i': i,
      'Hotel Name': row['Hotel Name'],
      'Review Stars': row['Review Stars'],
      'Reception': reception.sentiment.name,
      'Room': room.sentiment.name,
      'Service': service.sentiment.name,
```

Fuente: Elaboración propia

A continuación, recorremos las filas del documento y vamos almacenando las valoraciones en una variable *text*, para pasarla como parámetro al modelo, junto con la lista de aspectos que guardamos en la celda anterior (ilustración 28). Se pasa el conjunto de entradas al modelo preentrenado para hacer una predicción. Y, los resultados se los asignamos a variables que representen estas características para, posteriormente, averiguar su polaridad. A este bucle le hemos añadido una barra de progreso mediante la importación de la librería *tqdm*, puesto que la clasificación hace uso de muchos recursos y queríamos realizar un seguimiento.

El *pipeline* recoge no sólo las puntuaciones, sino también los estados ocultos, las de atención y sus gradientes con respecto a la salida del modelo. Al final, el método los empaqueta en el conjunto de salida y la devuelve.

Para averiguar la polaridad de cada característica dentro de un comentario, hacemos uso de *sentiment.name*, que nos da como resultado un valor único: *Positive* si se habla positivamente sobre el aspecto, *Negative* si es de forma negativa, o *Neutral* si en la valoración no se menciona el aspecto que estamos buscando o directamente expresa indiferencia ante ello.

Esta implementación se realizó en el cuaderno *ABSA.ipynb* y sus resultados se almacenaron en el archivo *absa_reviews_2019(GC).xlsx*.

6.4 Visualización con Power BI

Para el manejo de datos y creación de visualizaciones interactivas utilizaremos la aplicación de escritorio Power BI Desktop, que se puede descargar gratuitamente desde la tienda de Microsoft.

Lo primero que haremos será añadir los resultados que hemos obtenido en los pasos previos de desarrollo. Como ya hemos visto, estos datos pueden provenir de muchas fuentes diferentes; en nuestro caso los resultados derivados de las ejecuciones anteriores los hemos almacenados en tablas Excel. Por lo tanto, cargaremos los tres archivos: *absa_reviews_2019(GC).xlsx*, *hotels_info(GC).xlsx* y *translated_reviews_2019(GC)*.

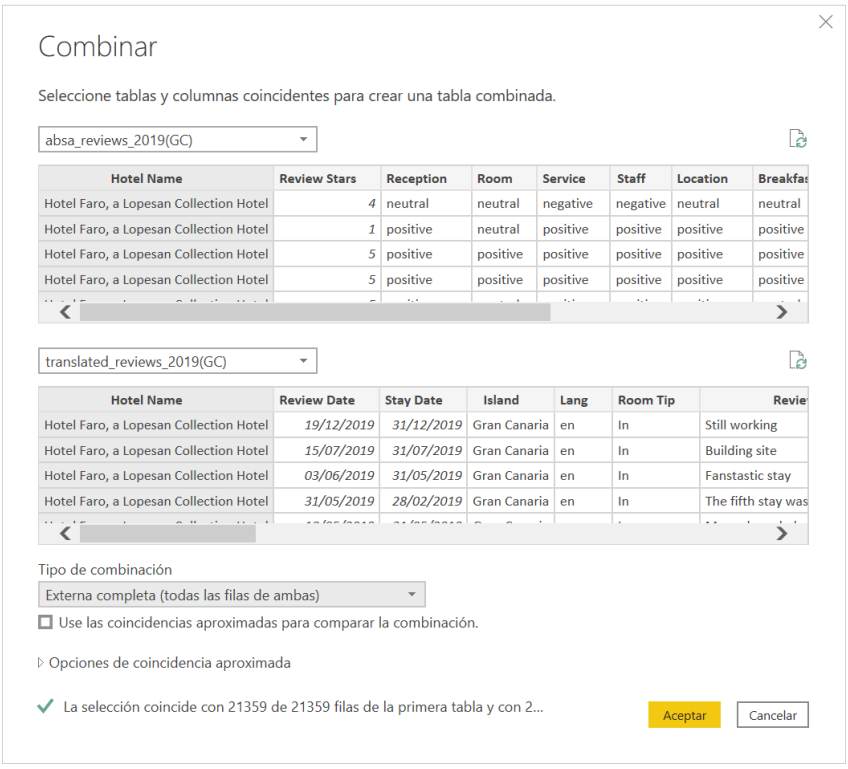
Por defecto Power BI detecta el tipo de datos que se almacena en cada columna, pero en nuestro caso tuvimos que especificar algunos de ellos como la categoría del hotel, el precio por noche o la URL, al no ser reconocidos por la aplicación.

Aparte de esto, limpiamos los valores referentes a las instalaciones, servicios de la habitación e idiomas hablados pues, al cogerlos directamente del script de la página, se almacenaron en formato de lista y había que prescindir de los corchetes y comillas simples que los englobaban. Además, añadimos dos nuevos campos en la tabla *hotels_info(GC)* correspondientes a la latitud y longitud de la ubicación de los diferentes alojamientos para una funcionalidad que veremos más adelante.

Con el propósito de que los datos quedaran más organizados y claros, quisimos separar los datos referentes a las reseñas y los relativos a las características de los hoteles. Para ello realizamos una combinación de consultas entre las tablas de *absa_reviews_2019(GC)* y *translated_reviews_2019(GC)*, tal y como podemos ver en la ilustración 29. El programa nos permite simplemente combinarlas o hacerlo añadiendo otra consulta. Escogimos esta última opción llamándola *reviews_2019(GC)*. Para poder realizar este paso, es necesario que ambas consultas originales tengan al menos una columna común que debemos seleccionar. En este caso se trataba de *Hotel Name*, el nombre del alojamiento.

Tras este paso, se podrán elegir aquellos campos de la segunda tabla (*translated_reviews_2019(GC)*) que queramos "expandir" y añadir a la nueva consulta. Hemos seleccionados aquellos relevantes para la visualización de los resultados (título de la reseña, la propia valoración, el tipo de viaje, y las estrellas de algunos aspectos). Es decir, que la nueva consulta se compondría de los campos de la tabla *absa_reviews_2019(GC)* más estos últimos.

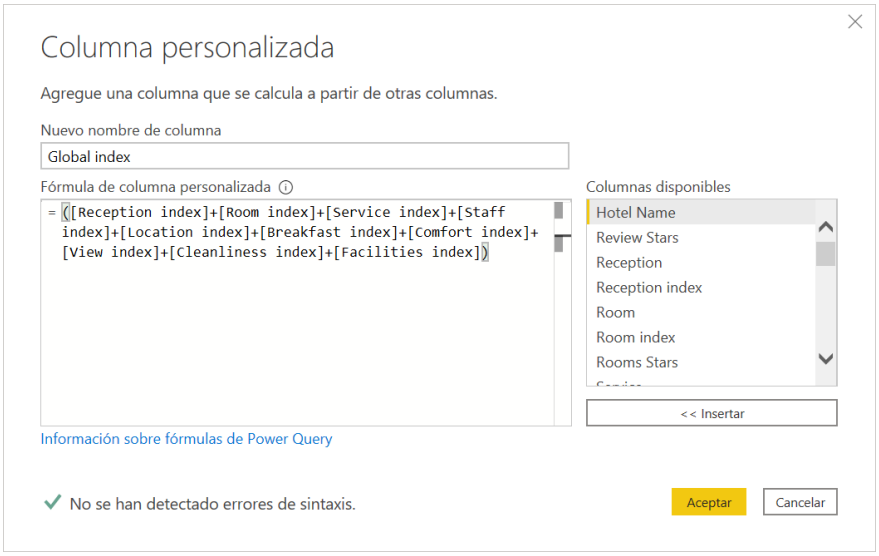
Ilustración 29. Combinación de consultas



Fuente: Elaboración propia

Por otro lado, en la nueva consulta, se consideró oportuno disponer de un campo que resumiera la polaridad del conjunto de los aspectos analizados. Para ello, se crearon columnas condicionales de cada uno de ellos, identificando el sentimiento con un valor numérico: para el positivo un 1; para el negativo un -1 y para el neutral un 0. Después, se estableció una columna personalizada que recogía la media de estos valores, como observamos a continuación:

Ilustración 30. Columna personalizada



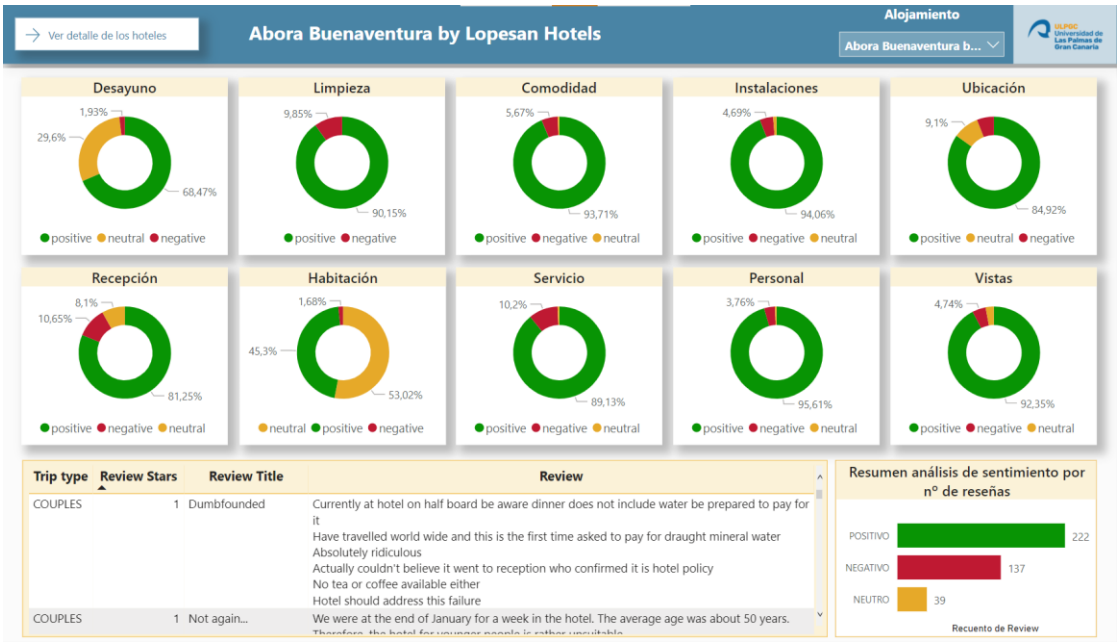
Fuente: Elaboración propia

Por último, agregamos un campo *Índice* mediante una nueva columna condicional que establecía la polaridad conjunta según el valor obtenido en la ecuación anterior. Si era mayor o igual que 0,05 representaba un sentimiento positivo, si era menor o igual a -0,05 representaba un valor negativo, y todo lo que estuviera en medio de esos valores, se consideraba un sentimiento neutral.

Una vez finalizadas las transformaciones de los datos, pasamos a la creación del panel y sus diferentes visualizaciones. Nos pareció oportuno dividir los resultados en dos pestañas: una con las resoluciones del análisis de sentimientos y la información sobre las reseñas, y otro con la información y características de los hoteles. Ambas con un tema personalizado según los colores del logo de la Universidad de Las Palmas de Gran Canaria.

La primera de las ventanas (ilustración 31) se compone de un menú desplegable en el que podemos seleccionar el alojamiento del que queremos visualizar los resultados del análisis. Estos datos se han elaborado a partir de un gráfico de anillos, en el que se representa el sentimiento de cada aspecto por colores: verde para una polaridad positiva, rojo para una negativa, y amarillo para una neutra. Encontramos diez gráficos como estos, correspondientes a las diez características que predefinimos en el análisis a nivel de aspecto, que muestran el porcentaje correspondiente a cada polaridad sobre el total. Al mismo tiempo, esta misma página incluye una tabla con las reseñas del hotel que esté seleccionado y sus características, además de un gráfico de barras que representa un resumen ABSA (con el campo *Índice*) según el número de valoraciones.

Ilustración 31. Visualización ABSA



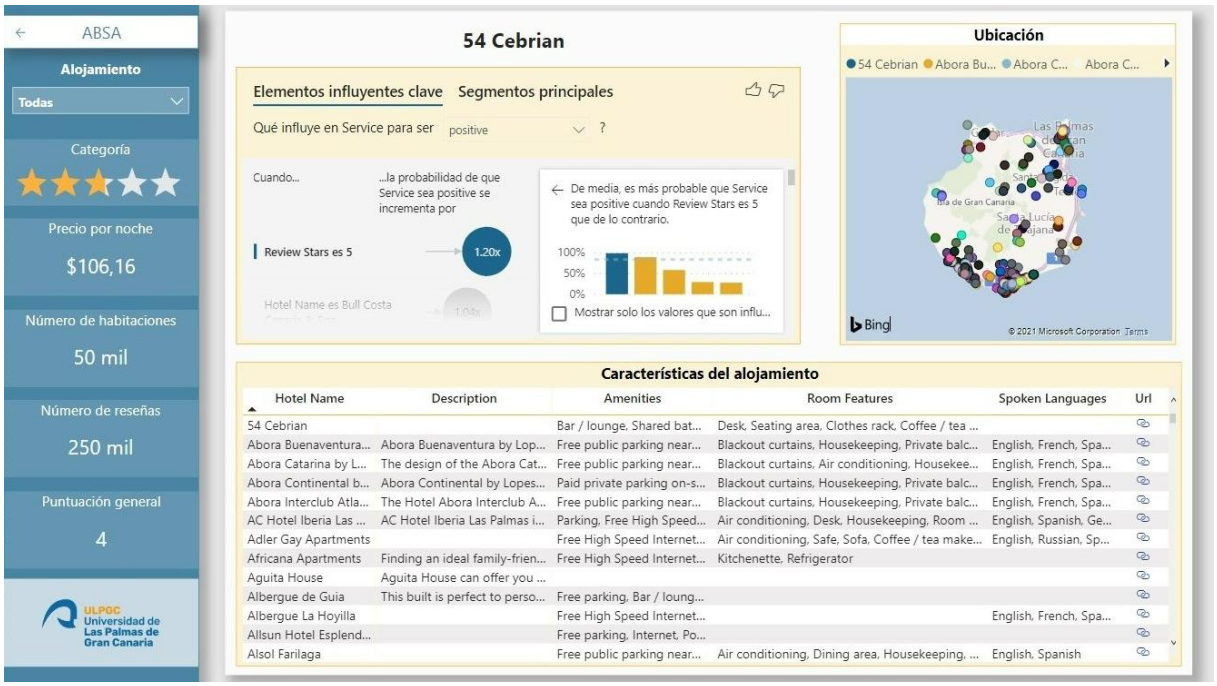
Fuente: Elaboración propia

No hay que olvidar que los paneles que creamos son interactivos, tanto para seleccionar el alojamiento como para ordenar las reseñas por cualquiera de sus campos. Esto se consigue clicando en cada uno de los encabezados de la tabla anterior.

Por otro lado, encontramos la segunda pestaña donde hemos incluido datos correspondientes a los alojamientos. En la parte izquierda del panel podemos observar algunas de las principales características que irán cambiando según el alojamiento que seleccionemos en el menú desplegable. Aparte de estas, incorporamos una tabla con más especificaciones del hotel, como son la descripción, las instalaciones, los servicios de la habitación, los idiomas o su URL. Esta tabla, al igual que la anterior, también es modificable en cuanto al orden en el que aparecen los datos.

Asimismo, otra de las ventajas que ofrece Power BI es la creación de mapas interactivos para los cuales necesitamos disponer de los mencionados campos latitud y longitud. Tal y como podemos ver en la ilustración 32, las localizaciones de los diferentes alojamientos se representan con una burbuja, y cada uno, con un color diferente. Sin embargo, si sólo seleccionamos uno de ellos, sólo aparecerá la ubicación del hotel que estamos estudiando.

Ilustración 32. Visualización detalles del hotel



Fuente: Elaboración propia

Por último, apreciamos uno de los visuales más novedosos que ofrece esta herramienta: los influyentes clave. Se trata de una de las muchas formas que tiene Power BI de incorporar la

inteligencia artificial, analizando una variable en función de otros campos. En este caso, queremos explicar cuándo toma una polaridad positiva el aspecto del servicio, según el nombre del hotel y la puntuación de las reseñas. También, en la pestaña que tenemos justo al lado llamada “Segmentos Principales”, encontramos aquellas combinaciones de valores de uno o más campos que incrementan la posibilidad de que el servicio sea positivo. Esto es sólo un ejemplo, pues consideramos que el servicio es uno de los aspectos más importantes y mencionados en una valoración, pero estos campos se pueden cambiar indistintamente.

Para la navegación entre páginas, se ha creado un botón que podemos identificar en ambas ilustraciones, en la parte superior izquierda del panel. Para su utilización, basta con presionar en el teclado del ordenador CTRL + hacer clic encima del recuadro.

Como ya hemos visto, los paneles son una forma excelente de mostrar una visión general sobre lo que está ocurriendo en el negocio. En este proyecto, los usuarios finales podrán disponer de una pantalla con los datos más relevantes como son los resultados del ABSA, objeto de este estudio, o los datos de los alojamientos. Además, cuenta con el gran añadido de poder interactuar con los diferentes gráficos.

7. RESULTADOS

En este capítulo se presentan los resultados obtenidos tras la ejecución del análisis, distinguiendo entre las diferentes implementaciones que hemos realizado a lo largo del trabajo. En primer lugar, se analizarán los datos derivados del *web scraping*, seguido de las traducciones de las reseñas de los turistas y terminando por analizar las valoraciones de los clientes de alojamientos turísticos con ABSA.

7.1 Experimento 1: Análisis del *web scraping*

El primer problema que nos encontramos a la hora de rastrear una web son los posibles bloqueos de la página debido al gran volumen de datos que se quiere almacenar y que se está analizando. Para ello hay que hacer que este proceso de scraping se parezca al de un humano navegando por la web, pues la mayoría de las tareas de rastreo tienen como objetivo hacerlo en el menor tiempo posible y la respuesta directa a esto es el bloqueo. En nuestro caso, incluimos un retardo aleatorio entre solicitudes, simulando esta navegación humana.

Por otro lado, nos dimos cuenta de que muchos datos relevantes para el estudio se obtenían de forma incorrecta mediante el uso de la librería Beautiful Soup debido a confusiones con los nombres de identificadores y clases. Esto lo resolvimos haciendo uso del ya explicado GraphQL, en el que pudimos obtener valores concretos y de forma conjunta en una sola consulta. Simplemente escribiendo el nombre de la variable o dato que nos interesara entre corchetes y comillas simples tras la respuesta a la petición, tal y como vimos en el capítulo anterior.

Este problema lo tuvimos no sólo en la obtención de los valores de las reseñas, sino en la recopilación de características de los alojamientos, para las cuáles hicimos uso de tres funciones adicionales con el fin de guardar el script de la página y tener acceso a datos "ocultos", que no se encontraban a simple vista en el sitio web.

A la hora de implementar el código, vimos la necesidad de incorporar una estructura *try/except* pues algunos datos no se podían almacenar y generaban errores en la ejecución. De esta manera, se probaría un bloque de código y si falla, continuaríamos con la siguiente iteración.

7.1.1 Resultados del *web scraping*

En la obtención de datos no hubo problemas pues todos se correspondían con los que se encontraban en la plataforma. En el caso del rastreo de las valoraciones, algunos de los comentarios disponían de mayores detalles de la estancia que otros, dependiendo de los datos introducidos por el turista a la hora de redactar el comentario.

Por el otro lado, y en referencia con las características de los alojamientos, algunos de ellos disponían de más información que otros, sobre todo los hoteles con respecto a los otros alojamientos (apartamentos, casas rurales, villas, fincas…). Y de estos, los de mayor categoría presentaban un conjunto de propiedades superior y más detallado de los servicios ofertados en su establecimiento.

Los resultados muestran que los alojamientos que más reseñas obtuvieron fueron hoteles de 4 y 5 estrellas y principalmente de las cadenas hoteleras Gloria Palace y Lopesan. Estos datos eran de esperar teniendo en cuenta que la lista la componen algunos de los hoteles más grandes de la isla de Gran Canaria.

Tabla 9. Los cinco hoteles con más reseñas

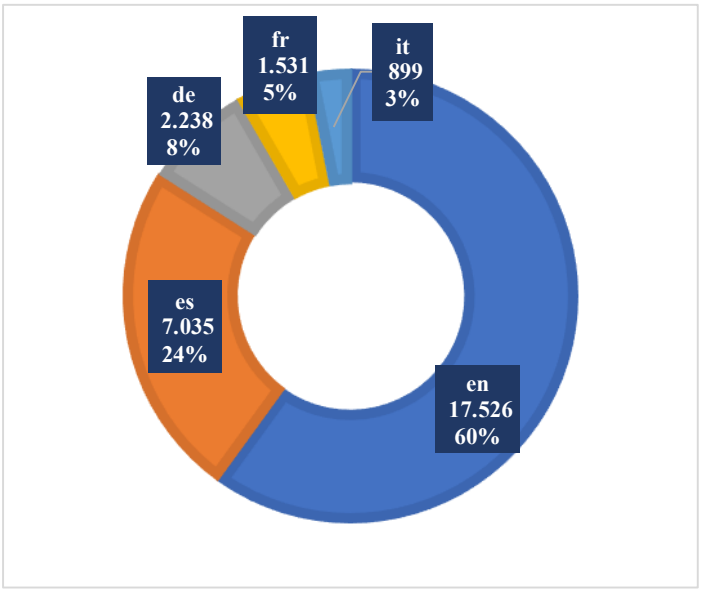
Alojamiento	***Reseñas***
Gloria Palace San Agustín Thalasso y Hotel	1177
Occidental Margaritas	901
Lopesan Costa Meloneras Resort, Spa y Casino	881
Gloria Palace Amadores Thalasso y Hotel	788
Lopesan Baobab Resort	756

Fuente: Elaboración propia

Por otro lado, encontramos que, de los idiomas seleccionados para el rastreo, el más utilizado fue el inglés con un 60% sobre el total, seguido del español y el alemán (ilustración 33). Esto concuerda con las conclusiones obtenidas en el análisis del capítulo cuatro, concretamente, al

apartado dedicado a la comprensión y preparación de los datos, donde se investigaban las nacionalidades que más visitaban las islas.

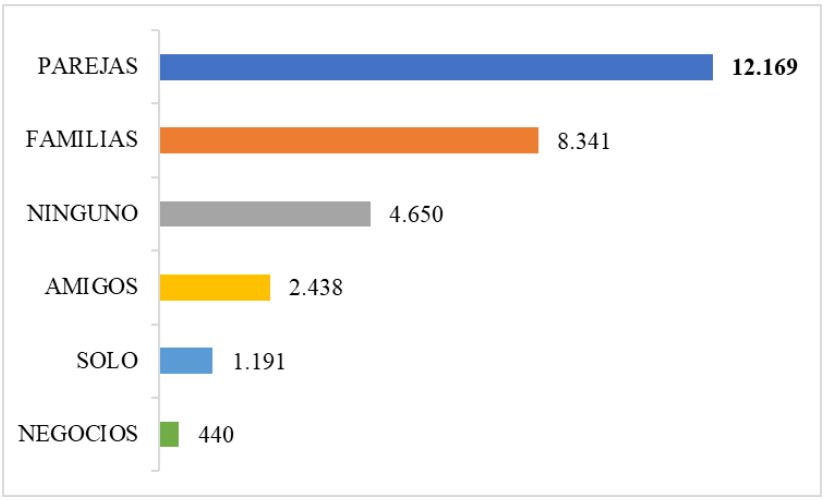
Ilustración 33. Distribución de idiomas en las reseñas



Fuente: Elaboración propia

Asimismo, y tras realizar un recuento sobre el tipo de viaje que realizaban los turistas, observamos que la mayoría de ellos viajaban a Gran Canaria con su pareja, seguido de las familias, tal y como observamos en la siguiente ilustración:

Ilustración 34. Tipo de viaje más popular entre los turistas

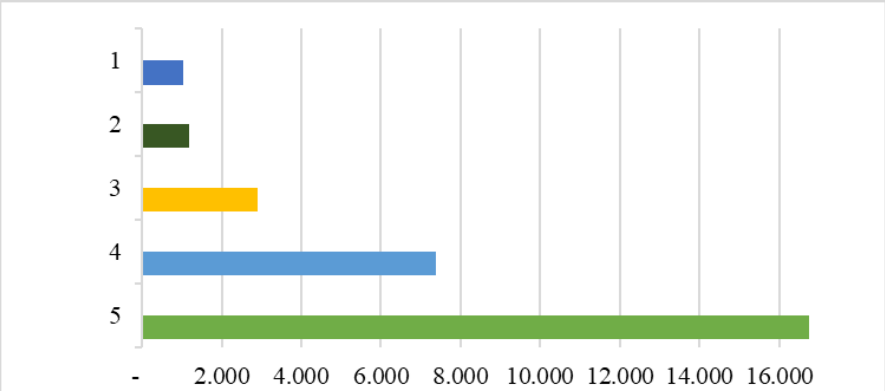


Fuente: Elaboración propia

Por otro lado, si tenemos en cuenta la calificación que los turistas asocian a las valoraciones, es decir, la conformidad y satisfacción que han tenido acerca de su estancia de forma general, vemos que, de las 29.229 reseñas totales, el 57% de los clientes las puntúa con un 5/5

(ilustración 35). Teniendo en cuenta este dato, es de esperar que, en nuestro posterior análisis de sentimientos, exista un porcentaje mayor de valoraciones que reflejen un sentimiento positivo hacia el alojamiento turístico que estemos estudiando.

Ilustración 35. Estrellas asociadas a las reseñas



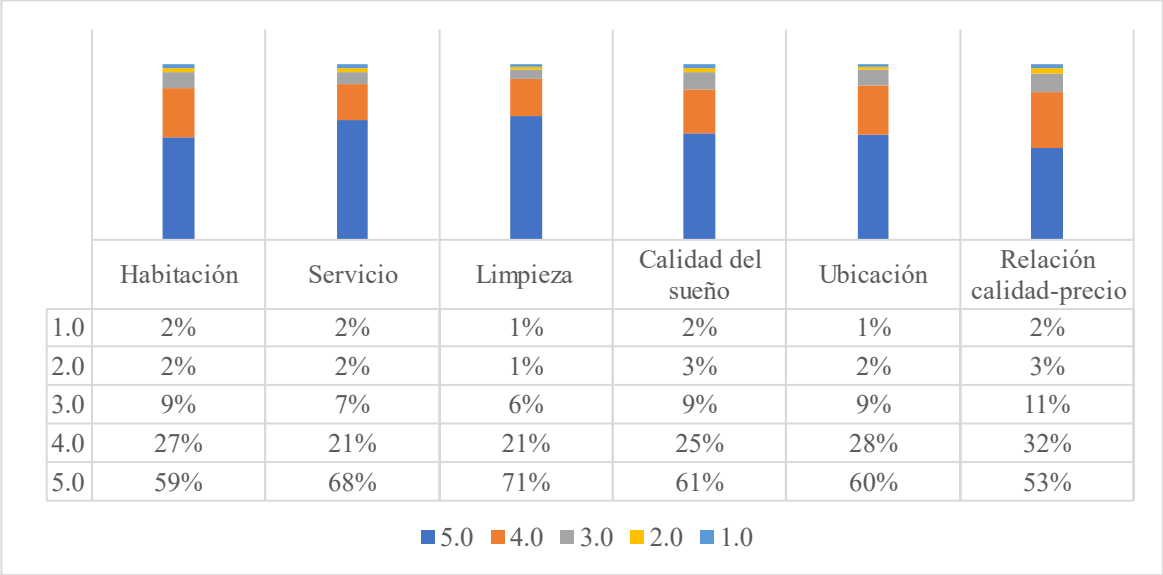
	5	4	3	2	1
Porcentaje	57%	25%	10%	4%	3%
Total	16.761	7.368	2.908	1.170	1.022

Fuente: Elaboración propia

Otra de las cuestiones que podemos intentar anticipar antes de realizar el análisis ABSA es la opinión que tienen los turistas sobre las diferentes características del alojamiento pues, recordemos del Marco Teórico, que las opiniones de TripAdvisor incluyen un sistema de valoración individual para algunos aspectos del establecimiento. En la ilustración 36 los podemos encontrar con sus correspondientes puntuaciones.

Tal y como observamos, todos ellos no fueron considerados para el posterior análisis de sentimientos, por lo que nos centraremos en los que sí tuvimos en consideración: la habitación, la limpieza, la ubicación y el servicio. Si contemplamos las puntuaciones de 1, 2 y 3 como negativas, observamos que las características que mayor porcentaje de reseñas acumulan son la habitación, la ubicación, el servicio y la limpieza, respectivamente. Consecuentemente, podríamos predecir que estos serán los aspectos peor valorados por los turistas tras el análisis que realizaremos en las próximas etapas.

Ilustración 36. Aspectos menos valorados según puntuación TripAdvisor



	Habitación	Servicio	Limpieza	Calidad del sueño	Ubicación	Relación calidad-precio
1.0	2%	2%	1%	2%	1%	2%
2.0	2%	2%	1%	3%	2%	3%
3.0	9%	7%	6%	9%	9%	11%
4.0	27%	21%	21%	25%	28%	32%
5.0	59%	68%	71%	61%	60%	53%

Fuente: Elaboración propia

Por otro lado, se reafirma la importancia que tiene responder a los comentarios de los clientes por parte de los responsables de los alojamientos turísticos con el fin de generar vínculos e incrementar la interacción con los turistas hospedados. De todas las valoraciones obtenidas en el proceso de *web scraping*, 20.076 obtuvieron una respuesta por parte de los establecimientos.

Tabla 10. Los diez hoteles con más respuestas

Alojamiento	*Reseñas totales*	*Respuestas del hotel*	*Porcentaje de respuestas*
Gloria Palace San Agustín Thalasso y Hotel	1176	1176	100,00%
Gloria Palace Amadores Thalasso y Hotel	788	788	100,00%
Lopesan Baobab Resort	756	755	99,87%
Radisson Blu Resort And Spa - Gran Canaria Mogan	686	598	87,17%
Abora Continental by Lopesan Hotels	537	534	99,44%
Abora Catarina by Lopesan Hotels	475	470	98,95%
H10 Playa Meloneras Palace	469	469	100,00%
Club Maspalomas Suites y Spa	421	419	99,52%
Meliá Tamarindos	424	406	95,75%
LABRANDA Playa Bonita	404	404	100,00%

Fuente: Elaboración propia

Como podemos observar en la tabla 10, las reseñas que más respuestas obtuvieron fueron aquellas provenientes de hoteles catalogados con 4 y 5 estrellas. Además, la mayoría de estos contestaron a la totalidad de reseñas obtenidas durante el año de estudio, lo que demuestra una preocupación por ofrecer el mejor de los servicios durante la hospedaje del turista. Cabe resaltar, que la mayoría de estas réplicas se realizaron en el mismo idioma que la valoración realizada por el turista, tal y como podemos apreciar en la tabla que tenemos a continuación.

Tabla 11. Ejemplo de respuesta a una valoración

Alojamiento	*Idioma*	*Título de la reseña*	*Reseña*	*Respuesta del alojamiento*
Hotel Faro, a Lopesan Collection Hotel	fr	Situation exceptionnelle	Un hôtel à la situation exceptionnelle, position stratégique: près de la plage et un 1/4 heure à pied de la plage naturiste, près du centre attractif des boutiques, bars ou restaurants. L'accueil à été parfait. Chambre au 5ème étage avec vue océan (1541), malheureusement trop bruyante du fait des aller-retour bus, voitures, camions d'approvisionnement et taxis. Point négatif, il n'y a pas de parking ou presque pas (nous avions pris une voiture de location). Grand point positif, le buffet: personnel accueillant et charmant. Les plats sont amples et divers tous les soirs avec différentes types de cuisine. Le petit déjeuner est très copieux. La salle est ample et bien positionnée devant l'océan. Un beau 4 étoiles à recommander.	Monsieur EricDuRocher, Un grand merci pour votre commentaire et merci de nous avoir fait part de votre expérience. Nous sommes ravis que vous ayez apprécié votre séjour parmi nous, Nous vous prions sincèrement de nous excuser pour les inconvénients produits pendant le bruit et nous regrettons beaucoup les dérangements J'espère vous pouvoir donner bientôt la bienvenue de nouveau. Cordialement, Relations Publiques

Fuente: Elaboración propia

Esto refleja una muestra de cercanía y confianza, en el que se crean relaciones entre gestores y consumidores. Se habla entonces de la comunicación bidireccional mencionada en el capítulo dos, donde los turistas sienten que sus propuestas y quejas están siendo escuchadas y tenidas en cuenta por parte de la organización. El buen trato con el cliente es importante tanto durante el periodo de estancia como después de este, de manera que garanticemos la máxima satisfacción del consumidor, objetivo de toda empresa dedicada al turismo. Además, se incita a este a recomendar los servicios ofertados y mostrar lealtad y compromiso con la marca, así como la repetición de compra.

7.2 Experimento 2: Análisis de las traducciones

A la hora de traducir las valoraciones nos encontramos con el problema de especificar la última versión de la librearía *googletrans* instalada, lo que resultó en una serie de errores constantes y en la imposibilidad de ejecutar del código. Descubierto el error, actualizamos la librería con la versión 4.0.0, pudiendo inicializar el *Translator* y analizar las valoraciones para la traducción.

Dentro del bucle observamos el mismo inconveniente que en el apartado anterior que resolvimos de la misma manera, mediante la introducción de un retardo aleatorio para no colapsar la peticiones a la API de Google y la implementación de una estructura *try/except*, ya que muchas de las reseñas no se podían traducir debido a varios factores, principalmente por la utilización de caracteres raros.

7.2.1 Resultados de las traducciones

Como ya vimos en el epígrafe dedicado a los Recursos Tecnológicos, *Google Translate* se considera uno de los traductores más utilizados y fiables hoy en día, admitiendo más de 100 idiomas, perfecto para esta parte del desarrollo del proyecto.

Para comprobar que efectivamente la API de *googletrans* ha funcionado correctamente, hemos seleccionado una muestra de 50 reseñas con el propósito de realizar una comparativa entre la valoración original y la final cuyo idioma destino era el inglés. Mostraremos un ejemplo con los 4 idiomas adicionales utilizados en el estudio: español, alemán, italiano y francés (tabla 12). Además, hemos comprobado la traducción con el segundo traductor más popular que encontramos, *Microsoft Translator*.

Tabla 12. Comprobación de la traducción de valoraciones

Alojamiento	Idioma	Título original	Reseña original	Título traducido	Reseña traducida
AxelBeach Maspalomas	de	Wunderbar	Toller Ort als Ausgangspunkt für eine gute Zeit in Maspalomas. Die Anlage ist gepflegt, das Frühstück reichhaltig, Pool und Wellness top. Die Mitarbeiter sind extrem freundlich, Danke Iván und Nereida	Wonderful	Great place as a starting point for a good time in Maspalomas. The complex is well maintained, the breakfast rich, pool and wellness top. The staff are extremely friendly, thanks Iván and Nereida
Lemon y Soul Las Palmas	es	.	Excelente muy buena ubicación, acogedor, limpio y sobre todo el responsable de los desayuno muy atento y educado me sorprendió gratamente.La recepción y camarera de piso muy atentas. Hotel recomendado. Repetiré	.	Excellent very good location, cozy, clean and above all the person responsible for the breakfast very attentive and polite surprised me pleasantly. The reception and very attentive floor waitress. Hotel recommended. I will repeat
Santa Monica Suites Hotel	it	Delusione	Non è un 4stelle superiore! Per € 300,00 a notte non sono assolutamente all'altezza! Non hanno letti matrimoniali, la colazione è scarsissima, la cena è deludente e fanno pagare il vino come se fosse del Barolo! Le lenzuola alcune volte sono sfilacciate L'unica convenienza sono le bellissime dune ad un passo! Evitatelo se potete	Disappointment	It's not a superior 4stelle!For € 300.00 per night I'm absolutely not up to it! They have no double beds, the breakfast is very poor, dinner is disappointing and charging the wine as if it were Barolo! The sheets sometimes are frayed the only convenience are the beautiful dunes to a step! Avoid it if you can
Hotel Silken Saaj Las Palmas	fr	Au top	Accueil chaleureux, souriant. Chambre parfaite, propre, agréable. Restau parfait. Un hôtel 4 étoiles qui le mérite amplement. Si j'ai l'occasion de retourner à Gran Canaria, j'y retournerai c'est certain.	In the top	Warm welcome, smiling. Perfect room, clean, nice. Perfect restau. A 4-star hotel that deserves it.If I have the opportunity to return to Gran Canaria, I will return it's certain.

Fuente: Elaboración propia

Los resultados muestran que la mayoría de las traducciones se realizaron de forma correcta, a excepción de aquellas palabras mal escritas o abreviaturas que utilizan algunos turistas. Por ejemplo, en la reseña italiana, el cliente no separó "4stelle" y el traductor lo mostró tal cual cuando debería haberlo traducido a "4 estrellas". O, en la reseña francesa, no se detectó la abreviación "Restau" cuando se refería al restaurante.

La utilización de esta API implica que, independientemente del idioma en el que se encuentre la valoración, podemos obtener una uniformidad en los datos traduciéndolas todas a un idioma común y de forma fiable.

7.3 Experimento 3: Análisis ABSA

Para este último paso en la experimentación hemos escogido el mismo grupo de datos que en la fase de traducción, de manera que resulte más sencillo examinar y estudiar los resultados obtenidos.

Lo primero que detectamos es una limitación en el número de caracteres a examinar, resultando en una disminución en el número de reseñas finales con respecto a las que teníamos originalmente. Esto sucede debido a que el modelo utilizado BERT tiene un límite de 512 tokens de entrada. La consecuencia directa en los resultados fue que las valoraciones se redujeron de las 50 que teníamos inicialmente a 42. Es decir, hemos tenido que prescindir de ocho valoraciones a la hora de realizar el análisis a nivel de aspecto. Aun así, la cantidad de datos siguió siendo bastante significativa con respecto a los iniciales como para poder sacar firmes conclusiones.

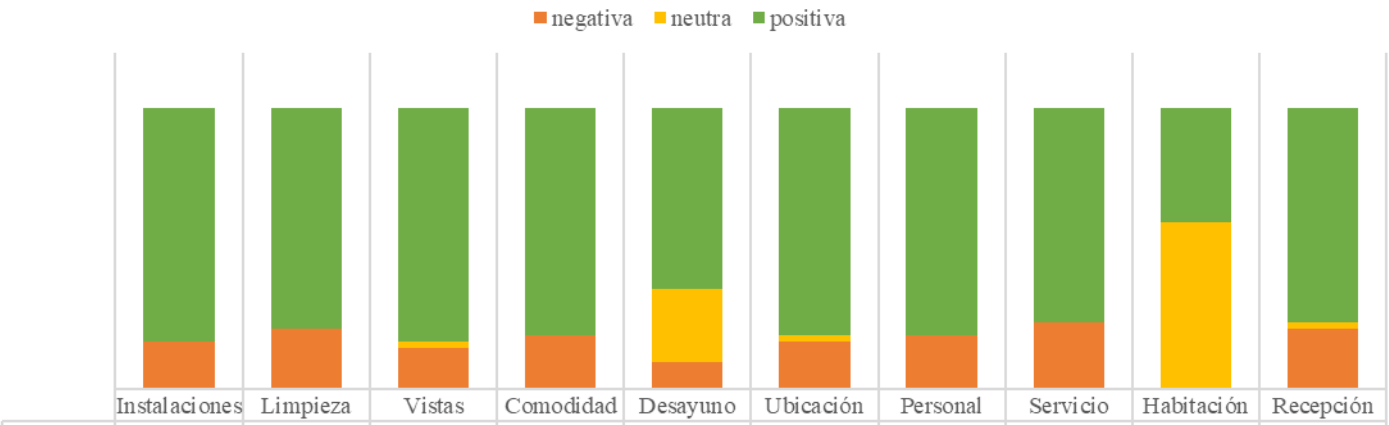
7.3.1 Resultados ABSA

De los diez aspectos analizados (recordemos que eran la recepción, la habitación, el servicio, el personal, la localización, el desayuno, la comodidad, las vistas, la limpieza y las instalaciones) todos mostraban un sentimiento positivo en la mayoría de las reseñas, a excepción de la habitación, en el que predominaba una polaridad neutral, tal y como observamos en la ilustración 37. Este resultado nos lo podíamos esperar del estudio que hicimos en fases anteriores, donde la puntuación que más daban los turistas era de 5/5, traduciéndose en un sentimiento positivo general.

Recordemos que una polaridad positiva la encontramos cuando se expresa una opinión de satisfacción hacia el alojamiento; una negativa cuando no se ha disfrutado de la estancia y se critica el servicio ofertado; y una neutra cuando no se menciona el aspecto estudiado o existe

un sentimiento tanto positivo como negativo hacia la característica y, por lo tanto, se contrarrestan produciendo la neutralidad.

Ilustración 37. Distribución ABSA en la muestra escogida de 50 valoraciones



	Instalaciones	Limpieza	Vistas	Comodidad	Desayuno	Ubicación	Personal	Servicio	Habitación	Recepción
positiva	83,33%	78,57%	83,33%	80,95%	64,29%	80,95%	80,95%	76,19%	40,48%	76,19%
neutra	0,00%	0,00%	2,38%	0,00%	26,19%	2,38%	0,00%	0,00%	59,52%	2,38%
negativa	16,67%	21,43%	14,29%	19,05%	9,52%	16,67%	19,05%	23,81%	0,00%	21,43%

Fuente: Elaboración propia

De acuerdo con la distribución anterior podemos confirmar que los aspectos mejor valorados por los turistas son las instalaciones y las vistas, acumulando un 83,33% de las reseñas cada uno. Por otro lado, los aspectos que acumulan más valoraciones negativas son el servicio, la recepción y la limpieza, con un 23,81% para el primero y un 21,43% para los dos últimos.

Si tenemos en cuenta las predicciones que hicimos al principio de la experimentación, en el análisis de *web scraping*, se confirma la conjetura de que el servicio y la limpieza serían los aspectos peor valorados por los clientes de los alojamientos turísticos. Recordemos del capítulo cuatro que son dos de las características que más valoraban los turistas a la hora de escoger y alojarse en un hotel, por lo que los establecimientos turísticos que hayan recibido estas críticas deben prestarles especial atención si quieren atraer a consumidores potenciales.

No obstante, corroboraremos el funcionamiento del algoritmo y resolución de los datos conseguidos con varios ejemplos. Recordemos que todos se encuentran en inglés porque es el idioma utilizado por el programa para realizar el análisis de sentimientos a nivel de aspecto.

Tabla 13. Ejemplo de valoración positiva

Alojamiento	*Puntuación*	*Reseña*
Hotel Riosol	4	Hotel great, friendly staff fantastic location for shops bars and quality restaurants. The balconies are massive with breath-taking views. Room - big enough and clean, shower room smelt pretty bad if the door was left open by the maid, then when we returned the whole apartment stank. would go back to this hotel. Pool and bars were clean and great. there was entertainment on most nights, but we went out so didn`t see it. Also, we were self-catering so did not eat at the hotel either.

Fuente: Elaboración propia

Como podemos ver en la tabla 13, al principio de la reseña menciona que el personal es amable y la ubicación fantástica. El cliente valora muy positivamente las impresionantes vistas que tiene el hotel. Sobre la habitación menciona que era grande y estaba limpia, pero más adelante se queja del olor que desprendía la ducha y se repartía por toda la habitación. Seguidamente, señala que tanto la piscina como los bares estaban limpios y eran geniales, además de tener entretenimiento por las noches. Por último, indica que no tuvieron la oportunidad de probar la comida del hotel.

Con esta información, la ejecución del algoritmo ha dado lugar a los siguientes resultados:

Tabla 14. Resolución ABSA de valoración positiva

Recepción	**Habitación**	**Servicio**	**Personal**	**Ubicación**	**Desayuno**	**Comodidad**	**Vistas**	**Limpieza**	**Instalaciones**
positiva	neutral	positiva	positiva	positiva	neutral	positiva	positiva	positiva	positiva

Fuente: Elaboración propia

Tal y como se esperaba, para los aspectos mencionados al comienzo de la reseña (personal, ubicación y vistas) la polaridad resultante es positiva. Cuando describe la habitación, indica que es grande, signo de una comodidad con sentimiento positivo; y que está limpia, indicando una polaridad positiva de la habitación. Sin embargo, critica el olor que desprende la ducha del cuarto, lo que se traduce en un sentimiento negativo. Por lo tanto, el sentimiento positivo del principio se contrarresta con esta opinión, resultando en una polaridad neutral hacia el aspecto de la habitación. También indica que la piscina y bares estaban limpios, explicando la polaridad positiva de la limpieza e instalaciones. A continuación, menciona favorablemente el entretenimiento ofrecido por el hotel, resultando en un sentimiento positivo hacia el servicio. Como apunta al final de la reseña, no comieron en el hotel por lo que no se puede valorar el aspecto relacionado con ello, en este caso, el desayuno. Por ese motivo la polaridad resultante es neutral. Y, a pesar de no mencionar la recepción explícitamente, sí indica que el personal fue

amable, por lo que la aplicación puede haber otorgado un sentimiento positivo a este aspecto por esa razón.

Tabla 15. Ejemplo de valoración negativa

Alojamiento	*Puntuación*	*Reseña*
NH Imperial Playa	2	This property has a much lower quality from the NH chain standards. Based his business only on his location, which is excellent and on the restaurant room that has a beautiful view. For the rest the structure is tired: - Old windows that make any noise pass - crumbling baths, with taps that fail to mix hot water coming out boiling or cold; even the bathroom toothbrush is missing, - There's a non-good smell around - Room services are missing, as you can make a coffee or a tea. - The floors are in fake marble, frozen !!! And they don't put mats or slippers. The comparative price with new structures is not even so convenient. Of good there is that they can improve with great ease if they decide to take it to the standard of other hotels they have. Of "Imperial" has absolutely nothing !!!

Fuente: Elaboración propia

En la tabla 15 encontramos otro ejemplo donde, desde el principio de la reseña, el turista reclama que el hotel en el que se hospedó tiene mucha menos calidad que el resto de los hoteles de la cadena, indicando que seguramente se tratará de una valoración negativa. Simplemente señala dos aspectos como positivos: la ubicación y la preciosa vista desde el restaurante. Después, menciona aspectos negativos de las infraestructuras del hotel, como baños deteriorados, ventanas viejas y rotas, o suelos de falso mármol. Además, demanda algunos servicios de la habitación como café o té, y se queja del olor que hay en el ambiente.

Con esta información, la ejecución del algoritmo ha dado lugar a los siguientes resultados:

Tabla 16. Resolución ABSA de valoración negativa

Recepción	**Habitación**	**Servicio**	**Personal**	**Ubicación**	**Desayuno**	**Comodidad**	**Vistas**	**Limpieza**	**Instalaciones**
negativa	neutral	negativa	negativa	positiva	neutral	negativa	positiva	negativa	negativa

Fuente: Elaboración propia

Comenzamos por identificar la polaridad positiva de los aspectos que menciona al principio de la reseña: la ubicación y las vistas. Por otro lado, era de esperar que el algoritmo calificara como negativo el aspecto de las instalaciones, pues el turista habla de ventajas viejas y rotas, baños deteriorados, suelo de mármol falso… Asimismo, critica el mal e incómodo olor que había en la atmósfera y la falta de servicios en la habitación, traduciéndose en una polaridad negativa para los aspectos de servicio, limpieza y comodidad. Como no menciona nada acerca del desayuno o la comida del hotel, el análisis clasifica el atributo del desayuno como neutro. No

obstante, pese a no nombrar nada acerca de la recepción o el personal del hotel, el algoritmo establece un sentimiento negativo para estos aspectos. Esto puede ser debido a la cantidad de actitudes negativas que tuvo el turista hacia el hotel, resultando en una calificación negativa general para aquellos aspectos no mencionados explícitamente. Sin embargo, deberíamos perfeccionar esta clasificación en el futuro si queremos obtener un análisis fiable y certero.

Tabla 17. Ejemplo de valoración con puntuación positiva, pero comentario negativo

Alojamiento	*Puntuación*	*Reseña*
Abora Continental by Lopesan Hotels	5	Enjoyed our stay again. We have had about 6 previous holidays at this hotel. The only area that would improve our stay would better evening entertainment. Live music always seems to get the best reaction from guests in my experience. The evening entertainment on this visit was the worst I have ever experienced.

Fuente: Elaboración propia

Como podemos observar en la tabla 17, la puntuación que ha otorgado el turista a la reseña hecha es de un 5/5, por lo que antes de leerla podríamos intuir que todo su contenido será positivo hacia el hotel. Sin embargo, critica el entretenimiento nocturno ofrecido por el hotel, considerando que ha sido uno de los peores que ha experimentado. Por otro lado, menciona que han disfrutado de su estancia y que ya se habían hospedado en ese alojamiento en seis ocasiones anteriores, reforzando el efecto de la repetición de compra que mencionábamos en el Marco Teórico. Si estos turistas han vuelto al mismo hotel por séptima vez significa que están tan satisfechos con el trato recibido y los servicios ofertados, que se han convertido en clientes fieles y leales de este hotel en particular, el Abora Continental by Lopesan Hotels.

La ejecución del algoritmo ha dado lugar a los siguientes resultados:

Tabla 18. Resolución ABSA de valoración con puntuación positiva y comentario negativo

Recepción	**Habitación**	**Servicio**	**Personal**	**Ubicación**	**Desayuno**	**Comodidad**	**Vistas**	**Limpieza**	**Instalaciones**
positiva	positiva	negativa	positiva	positiva	neutral	positiva	positiva	positiva	positiva

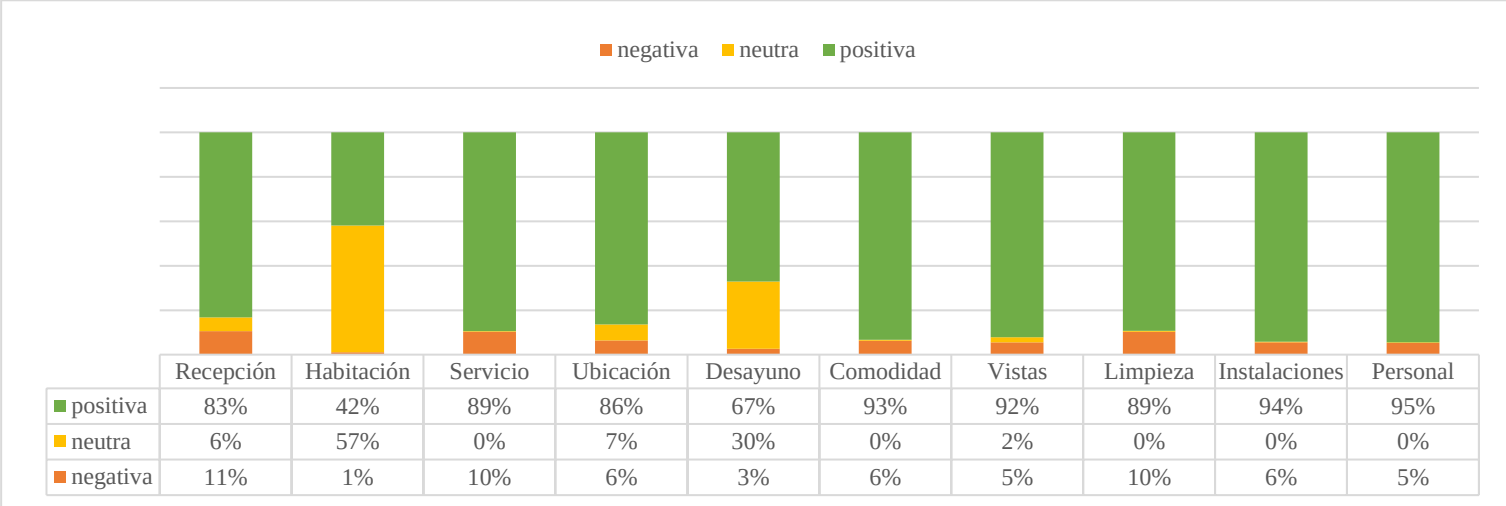
Fuente: Elaboración propia

Como el único comentario negativo se realizó hacia el entretenimiento nocturno, es decir, hacia los servicios ofrecidos por el hotel, era de esperar que este aspecto resultara en una polaridad negativa, tal y como observamos en la tabla 18. El turista no menciona explícitamente el resto de los aspectos, sino que simplemente comenta que disfrutó mucho de su estancia en el hotel. Seguramente el sistema haya detectado esta afirmación como un sentimiento positivo, otorgando a las características restantes una polaridad positiva. Sin embargo, esto no debería ser así, pues en ningún momento hace referencia a la recepción, habitación, personal… Por lo

que el programa debería haber clasificado esos aspectos como neutros, tal y como hizo con el desayuno. Nuevamente, si queremos perfeccionar los resultados del análisis, debemos realizar investigación y trabajo futuro en torno a este punto, con el fin de que la aplicación clasifique e identifique perfectamente cada uno de los atributos utilizados en el estudio.

Seguidamente, mostraremos los resultados globales de la totalidad de las reseñas extraídas de todos los alojamientos turísticos de Gran Canaria (ilustración 39). La consecuencia directa fue que estas se redujeron de las 29.233 que teníamos inicialmente a 21.359. Es decir, se han perdido 7.874 valores a la hora de realizar el análisis a nivel de aspecto. Aun así, la cantidad de datos siguió siendo bastante significativo como para poder sacar firmes conclusiones.

Ilustración 38. Distribución ABSA de la totalidad de reseñas



	Recepción	Habitación	Servicio	Ubicación	Desayuno	Comodidad	Vistas	Limpieza	Instalaciones	Personal
positiva	83%	42%	89%	86%	67%	93%	92%	89%	94%	95%
neutra	6%	57%	0%	7%	30%	0%	2%	0%	0%	0%
negativa	11%	1%	10%	6%	3%	6%	5%	10%	6%	5%

Fuente: Elaboración propia

En la ilustración 38 podemos encontrar la distribución que obtendríamos tras realizar el análisis de sentimientos a nivel de aspecto a la totalidad de reseñas de los alojamientos turísticos de Gran Canaria para el año 2019. Como podemos observar, los aspectos que más valoran los turistas son el personal, seguido de las instalaciones, dos de las características claves para prosperar en el sector servicios. De la misma forma, las vistas también son valoradas muy positivamente por los turistas, atributo esencial teniendo en cuenta la situación geográfica de los diferentes alojamientos en la isla.

Por otro lado, los atributos que peor valoran los viajeros, de manera general, son la recepción, la limpieza y el servicio, coincidiendo con la muestra de 50 alojamientos que seleccionamos al principio. Como vimos en apartados anteriores, los dos últimos se consideran una de las características que más importancia tienen para los turistas a la hora de escoger un alojamiento.

Es por ello por lo que los gestores hoteleros deben prestar especial atención a estos aspectos si quieren asegurar el bienestar y posterior fidelidad de sus clientes.

En definitiva, los resultados obtenidos durante el desarrollo de este experimento eran de esperar pues el análisis de sentimientos que estamos realizando es a un nivel muy profundo, complejo y del que se tiene poca información hoy en día. Como todo algoritmo de análisis, siempre existirá un porcentaje de error y más aún, considerando que la tecnología utilizada es tan novedosa que aún faltan estudios y tiempo para perfeccionarla.

No obstante, realizaremos un último estudio comparativo de los hoteles de Gran Canaria y Tenerife. En este caso, hemos hecho una selección de aquellos con categoría entre 4 y 5 estrellas, y con más de 500 habitaciones. Para los datos de la isla de Tenerife, hemos tenido que realizar todo el proceso de desarrollo del trabajo, comenzando por el *web scraping*, pasando por la traducción de las reseñas al inglés, y terminando con el ABSA. La lista de hoteles la podemos encontrar a continuación:

Tabla 19. Selección de hoteles de 4 y 5 estrellas con más de 500 habitaciones

Hoteles	***Reseñas***
Gran Canaria	**2253**
Lopesan Costa Meloneras Resort, Spa y Casino	733
Lopesan Baobab Resort	578
Lopesan Villa del Conde Resort y Thalasso	414
Hotel Riu Gran Canaria	328
Hotel Riu Palace Meloneras	200
Tenerife	**2131**
Gran Melia Palacio De Isora	902
Guayarmina Princess	497
Bahia Principe Sunlight Costa Adeje	400
Iberostar Bouganville Playa	332
Total general	**4384**

Fuente: Elaboración propia

Como era de esperar, aquellos alojamientos de mayor clase suelen tener mejores comentarios que los de categorías inferiores, es decir, comentarios más positivos que aquellos hoteles con menos estrellas. Esto es debido a que se consideran alojamientos de "lujo" y uno de los principales objetivos es ofrecer un servicio como tal, además de mantener una reputación online, imprescindible para sus resultados económicos. Teniendo esto en cuenta, observamos que todos los aspectos de los hoteles estudiados presentan mayoría de polaridad positiva, tal y como podemos ver en la siguiente ilustración.

Ilustración 39. Resumen ABSA de Gran Canaria y Tenerife

Fuente: Elaboración propia

Al comparar dos islas de características similares y alojamientos turísticos con las mismas propiedades, se podía predecir que tanto Gran Canaria como Tenerife coincidirían en los datos resultantes del análisis de sentimientos a nivel de aspecto, tal y como podemos ver en la figura 39. Por consiguiente, los aspectos que más valoran los turistas son las instalaciones, con un 95,47% en Gran Canaria, y un 93,71% en Tenerife; y el personal con un 95,07% en Gran Canaria y un 93,43% en Tenerife.

Por otro lado, los que menos valoran son la recepción, con un 10,56% en Gran Canaria y un 11,92% en Tenerife; el servicio, con un 10,25% en Gran Canaria y un 11,54% en Tenerife; y la limpieza, con un 10,87% en Gran Canaria y un 13,42% en Tenerife. En este último, la diferencia porcentual entre las islas es significativa en comparación con el resto de los aspectos, por lo que Tenerife debe hacer hincapié en la mejora de esta característica si quiere seguir siendo un rival de competencia directa no sólo para Gran Canaria, sino cualquier alojamiento turístico con propiedades y servicios similares. En términos generales, Tenerife acumula un número de reseñas negativas mayor que Gran Canaria, aunque esta diferencia no es alarmante dado el volumen de datos con el que estamos tratando.

No obstante, de forma conjunta, los resultados muestran un porcentaje muy alto de reseñas con polaridad positiva; consecuencia totalmente esperada, teniendo en cuenta las características de los hoteles que hemos seleccionado para esta parte del análisis.

Finalmente, cabe mencionar que, si queremos realizar un análisis exhaustivo de cada alojamiento en relación con cada aspecto seleccionado para el estudio, podremos encontrarlo en el panel generado con Power BI, cuyo enlace encontramos en el anexo.

8. USO DE LA HERRAMIENTA DESARROLLADA EN EL ENTORNO EMPRESARIAL

En este apartado plantearemos algunos ejemplos de cómo los alojamientos pueden beneficiarse de la utilización del modelo creado que engloba tanto la recopilación de las valoraciones de los turistas, como su traducción y posterior análisis de sentimientos.

Este sistema está pensado para los gestores hoteleros, pues son los encargados de administrar y gestionar la organización, siempre con el propósito de satisfacer a los clientes en referencia al servicio ofertado. Se encargan de valorar el negocio desde un punto de vista objetivo y con un tiempo de respuesta rápido ante los problemas que surjan, siempre con la intención de mantener la reputación del alojamiento y promoverlo ante futuros consumidores.

Con la incorporación de este prototipo a su negocio, el gestor hotelero tendrá la posibilidad de analizar y comparar grandes cantidades de datos acerca del pensamiento que tienen los turistas sobre su establecimiento. De esta manera, podrían replantear y tomar decisiones en cuanto a la administración y dirección que sigue la empresa, buscando nuevas oportunidades de negocio.

Los datos o resultados que se van a obtener tras la ejecución será el sentimiento que tienen los clientes tras su estancia. Y no sólo eso, sino una descomposición de los diez aspectos más

valorados en un alojamiento turístico: el desayuno, la limpieza, la comodidad, las instalaciones, la ubicación, la recepción, la habitación, el servicio, el personal y las vistas.

El producto de este análisis de sentimientos a nivel de aspecto lo podremos visualizar con la herramienta Power BI, diseñada para poder seleccionar cualquier alojamiento de cualquier lugar que nos interese y obtener, por un lado, el propio ABSA de las valoraciones de los clientes y, por el otro, la información detallada de un hotel en específico.

El programa dispone de un menú en que podemos seleccionar aquel o aquellos alojamientos que queremos estudiar, pudiendo filtrarlos por algunas de sus características como son el número de estrellas, habitaciones, tipo de alojamiento, zona en la que se encuentre… Una vez realizado eso, obtendremos la polaridad resultante de los correspondientes aspectos estudiados por separado, y el sentimiento global en función de la totalidad de reservas para ese hotel.

Asimismo, dispondremos de una lista con las reseñas originales y sus propiedades, de manera que el gestor pueda leerlas y averiguar la causa de la polaridad para cada aspecto. Esta estructura también se podrá filtrar por el atributo que el responsable hotelero desee: tipo de viaje, puntuación, usuario…

Por otro lado, y como ya mencionamos previamente, podremos interactuar con la información que nos ofrece el programa sobre un alojamiento en particular. Se incluyen sus principales características, como son su categoría, el número de habitaciones, el número de reseñas, la puntuación general otorgada por los turistas y un mapa interactivo donde visualizar su ubicación. Además de estas, se podrá visualizar el resto de las cualidades que definen al alojamiento como son su descripción, las instalaciones, los servicios de la habitación, los idiomas que hablan los empleados y el enlace a su página web.

Al mismo tiempo, hemos querido ir más allá con la incorporación de un apartado dedicado a la inteligencia artificial en donde se muestran los campos que actúan sobre la obtención de un valor concreto de un atributo. En nuestro ejemplo, se muestran los conjuntos que influyen en que el aspecto del servicio tenga una polaridad positiva. A este respecto, los campos seleccionados para su explicación han sido el nombre del hotel y las estrellas de las reseñas. No obstante, estos valores son fácilmente intercambiables dependiendo de la finalidad que tenga el alojamiento y en qué característica se quiera centrar.

Como vemos, no resulta nada complicado empezar a utilizar este servicio pues es muy intuitivo y accesible, en el que no se requieren conocimientos previos acerca de la programación que hay detrás ni de las librerías utilizadas durante el desarrollo del sistema.

Consecuentemente, supondría un cambio en el modelo de negocio de las empresas hoteleras que utilicen esta aplicación, pues generaría una propuesta de valor ventajosa frente a sus competidores.

A continuación, presentaremos cuatro casos de uso que podemos realizar teniendo como base el prototipo presentado en este trabajo, en el que explicaremos los cambios que tendremos que hacer tanto a nivel de código como visual.

8.1 Caso de estudio 1: Hoteles de Gran Canaria

Empezaremos definiendo uno de los casos más simples y con menos modificaciones que nos podemos encontrar, pues ya se encuentra prácticamente implementado en su totalidad:

Un directivo de un hotel de Gran Canaria quiere realizar un análisis de competencia con sus rivales directos, es decir, con aquellos hoteles que tengan las mismas características que el suyo. En este caso, se presupone que se trata de un alojamiento de tamaño medio (40 habitaciones) y categoría tres estrellas

A nivel de código no debemos modificar nada puesto que los competidores que queremos evaluar se encuentran en el mismo destino que nosotros, Gran Canaria. Por lo que los enlaces tanto del *web scraping* de las reseñas como de la información sobre los hoteles permanecen inalterables.

Pasaríamos entonces directamente a la visualización de Power BI donde, por un lado, tendremos una pestaña con la información sobre los aspectos más y menos valorados por los turistas y, por el otro, una pantalla con los detalles sobre cada uno de los alojamientos.

Si queremos conocer cuáles son los hoteles con nuestras mismas características y dónde se encuentran ubicados nos centraríamos en esta segunda parte, pues gracias a los filtros que ofrece el software de análisis empresarial, podremos indicar que muestre los resultados exclusivamente de aquellos alojamientos de tres estrellas y de entre veinte y sesenta habitaciones.

La utilización de estos filtros es muy fácil de usar: simplemente debemos arrastrar aquellos campos que queramos utilizar (en este caso serían la categoría del hotel y el número de

habitaciones) para filtrarlos por los valores que nos interesen. Para ello se puede realizar un filtrado básico o avanzado. En este caso, utilizaremos el primero para la categoría del alojamiento, y el segundo para poner el rango del número de habitaciones, tal y como observamos en la ilustración 40.

Este cambio deberemos hacerlo en ambas pestañas de tal modo que se actualice, automáticamente, el listado de hoteles que tenemos. De esta forma, el mapa interactivo también quedará reestablecido con las nuevas ubicaciones de aquellos alojamientos seleccionados, así como los aspectos a estudiar.

Ilustración 40. Caso de estudio 1

Fuente: Elaboración propia

Como podemos observar, el listado de hoteles que teníamos originalmente se ha reducido de forma considerable, permitiendo el estudio de nuestra competencia directa. Si lo que nos interesa es examinar los resultados de cada alojamiento de forma independiente, lo haremos mediante el uso del menú desplegable que tenemos en la parte izquierda de la imagen.

Este análisis generaría una gran ventaja competitiva frente al resto de alojamientos, pues nuestro hotel dispone de un software muy potente que nos permite valorar de forma automatizada, rápida y eficaz las preocupaciones, recomendaciones y opiniones que tienen nuestros clientes y los de nuestros competidores acerca de su estancia. De esta forma, buscaremos debilidades que debemos solventar, así como puntos fuertes que potenciar, en nuestros alojamientos y

servicios para ofrecer la mejor de las experiencias a nuestros turistas, garantizando la satisfacción de estos y consiguiendo un incremento en las ventas y reputación del hotel.

Esto resulta de gran importancia si tenemos en cuenta el contexto en el que nos encontramos, donde se están evaluando hoteles de Gran Canaria, uno de los sitios turísticos más visitados del panorama nacional y cuya economía se sustenta gracias a la industria del turismo. Por lo tanto, tenemos la obligación de brindar los mejores servicios y prestaciones a lo largo de toda la duración de la estancia de nuestros turistas.

8.2 Caso de estudio 2: Cadena hotelera

Para este segundo caso hemos elaborado un supuesto que nos obliga a realizar algunas modificaciones en cuanto al código que disponemos originalmente:

Los dueños de una gran cadena hotelera que tiene alojamientos distribuidos por todo el mundo desean realizar un análisis y comparación de los aspectos que más y menos valoran los clientes en cada zona y de cada hotel

Por consiguiente, comenzaremos por cambiar el enlace inicial del *web scraping*, tanto en el correspondiente a las reseñas como en el de las características de los hoteles. En su lugar, indicaremos la URL que englobe todos los hoteles de nuestra cadena, independientemente del destino en el que se encuentren.

Igualmente, como queremos incluir todas las nacionalidades de manera que podamos realizar un análisis más exhaustivo, eliminaremos la selección de idiomas que teníamos preestablecidas. Recordemos que se trataban de los idiomas inglés, alemán, francés, italiano y español.

En la visualización del Power BI tendríamos entonces un mapamundi con todas las localizaciones de nuestros hoteles, sus características y valoraciones. Dispondríamos de, por una parte, un panel con las propiedades de cada hotel de la cadena y, por la otra, una pestaña con todas las valoraciones de nuestro conjunto hotelero, así como los aspectos analizados en torno al sentimiento que tuvo el turista durante su estancia.

De la misma forma que en el caso anterior, filtraremos los resultados por aquellos parámetros que queramos analizar. En este caso, nos gustaría saber el sentimiento tanto general como por aspectos que tiene un conjunto de hoteles en una zona determinada. Para ello, arrastraremos el campo de la ubicación a la sección de filtrado y seleccionaremos aquel destino que nos interesa estudiar. La resolución será un panel con todos los hoteles localizados en el lugar escogido, que

podremos analizar de forma independiente con el uso del menú desplegable mencionado en ocasiones anteriores.

Por ende, sabremos qué aspectos son los que valoran los turistas a nivel nacional y, por lo tanto, ajustar sus necesidades y requerimientos, adaptando cada hotel al lugar en el que se encuentra ubicado. Tendremos que buscar entonces una estrategia de segmentación geográfica para diferenciar el servicio ofertado según la localización, consiguiendo una diversificación y expansión de nuevos mercados (Rodrigo García, 2014), siempre teniendo en cuenta las características del entorno, del propio establecimiento y los criterios de la demanda.

Los resultados de este estudio nos aportarían mucha información relevante acerca del posicionamiento global que tiene nuestra cadena hotelera y es nuestra labor comprender lo que está ocurriendo con los diferentes establecimientos que tenemos para aportar soluciones de mejora y adaptados a cada zona. Porque, como ya hemos visto a lo largo del trabajo, la visión que tienen nuestros clientes acerca de nuestro hotel constituye un factor determinante en la elección de un destino vacacional por parte de los futuros consumidores potenciales.

Es por esta razón por la que debemos adecuar nuestras instalaciones y servicios si queremos incrementar nuestra ventaja competitiva, además de nuestra demanda y oferta, incrementando las ventas e ingresos de nuestra empresa. Debemos pensar, no sólo la zona en la que se encuentra cada hotel, sino en la cadena hotelera en su conjunto, con el fin de garantizar una armonía, uniformidad y calidad en todos y cada uno de nuestros alojamientos.

8.3 Caso de estudio 3: Gestión hotelera global

El tercer ejemplo se centra en analizar las reseñas de los turistas en los diferentes alojamientos turísticos a nivel global:

Un gestor de una organización de gestión de un destino turístico desea comparar globalmente las valoraciones de los hoteles de Gran Canaria con otros destinos de competencia directa

Este es uno de los casos más complicados que nos podemos encontrar en términos de la limitación de recursos y el gran volumen de datos con el que se pretende trabajar. Para poder llevarlo a cabo, necesitamos que antes se mejore la escalabilidad del modelo y se dispongan de los medios necesarios para la descarga y análisis.

En primer lugar, debemos asegurarnos de cuáles son los destinos con los que vamos a competir de forma directa. A tal efecto, tendremos que analizar diferentes aspectos y criterios para definir

aquellos destinos similares, en términos turísticos, a la isla de Gran Canaria. Estos indicadores incluyen tanto el entorno económico, sociocultural, como medioambiental.

Cada vez son más los lugares que compiten con Canarias, principalmente por la incorporación de la actividad turística en muchos países donde antes se encontraba limitada por razones políticas, culturales o económicas; el incremento de viajes intercontinentales; los cambios en las prácticas de los consumidores, que son más propensos a elegir destinos extraordinarios; y el crecimiento de la industria del turismo tanto geográficamente hablando, como la generación de nuevos modelos de negocio.

El Plan de Marketing Estratégico 2018-2022 de Canarias identifica entonces los siete factores que más influyen a la hora de comparar un destino turístico con otro: la tipología de la oferta, la distancia entre lugar de origen y destino, el precio, la estacionalidad, la madurez del destino turístico, el marco jurídico y social (seguridad, garantías sanitarias…), y el grado de diferenciación y exotismo (Promotur Turismo de Canarias y Gobierno de Canarias, 2018).

Teniendo estos elementos en cuenta, podemos diferenciar y seleccionar varios grupos de sitios turísticos por los que muchos turistas reemplazarían Gran Canaria como su destino vacacional, pues ofrecen una propuesta de valor muy similar al de la isla. Grecia, Portugal, Chipre, Malta, o Baleares son algunos ejemplos de competidores directos.

En términos del software empresarial creado, debemos tener en cuenta todos estos destinos en la primera etapa, es decir, en la descarga de datos de la plataforma de viajes (*web scraping*). La forma más sencilla de hacerlo sería elaborando una lista con todos los destinos derivados del previo análisis de competencia que hemos realizado y, en el código, filtrar la descarga de los valores por ella. Además, el enlace inicial vuelve a cambiar por uno que englobe todos los alojamientos turísticos de todo el mundo.

Por otro lado, y teniendo en cuenta la gran cantidad de datos disponible, deberíamos establecer un periodo para las reseñas que vayamos a estudiar. Por ejemplo, obtener aquellas valoraciones realizadas en los últimos cinco o diez años, dependiendo de la capacidad y limitaciones de nuestros recursos. Así, de cara a los resultados finales, podríamos observar la evolución de las opiniones que han tenido los turistas en los diferentes destinos a lo largo del tiempo.

Además, si nos queremos centrar en la competitividad y los factores nombrados anteriormente, en la parte relacionada con el análisis ABSA, podríamos cambiar los aspectos existentes por

estos atributos, comprobando si los turistas valoran positiva o negativamente estas propiedades a la hora de escoger un destino turístico.

Los resultados los visualizaríamos en un gran cuadro de mandos que podría servir como fuente de información a muchas entidades, incluyendo los propios alojamientos de Gran Canaria, con el fin de buscar alternativas y estrategias con las que puedan adaptar sus modelos de negocio ante la amenaza de estos competidores.

8.4 Caso de estudio 4: Impacto de la COVID-19

Este último caso busca encontrar el impacto y la adaptación de los diferentes hoteles a la grave situación de crisis sanitaria que estamos viviendo:

Identificar cómo se han adaptado los diferentes alojamientos turísticos ante la pandemia de la COVID-19 y cómo han recibido los turistas estos cambios

Como ya sabemos, el coronavirus ha causado grandes estragos, afectando a muchas áreas, sobre todo a la industria del turismo. Ha supuesto un antes y un después en este sector, con un drástico descenso en el número de turistas, poniendo en riesgo el empleo de muchas personas del negocio y concluyendo en el cierre de muchos alojamientos.

Con este ejemplo se pretende observar qué alojamientos y de qué forma se han adaptado a los efectos de la crisis, cuáles no han podido salir adelante, y, de los pocos turistas que hayan podido alojarse en sus establecimientos, qué opinión tienen los clientes ante estos cambios.

Lo primero que haremos será filtrar por el año en el que empezó a afectar la pandemia, es decir, cogeremos un periodo de tiempo de aproximadamente dos años (desde finales de 2019 hasta ahora) y tendremos en cuenta todos los idiomas pues, al haber menos turistas, la información disponible se reduce considerablemente, así que consideraremos todas las nacionalidades.

Posteriormente, a la hora de realizar el análisis de sentimientos a nivel de aspecto, podríamos cambiar estos últimos por aquellos atributos más relevantes teniendo en cuenta el contexto en el que nos encontramos. Es decir, analizar las valoraciones de los turistas en términos de cómo se ha adaptado el alojamiento a esta adversidad. Por ejemplo, podríamos buscar términos relacionados con pandemia, virus, desinfectante, limpieza, distancia, seguridad, control…

Esto generaría una serie de resultados en los que se podría observar cómo tanto el número de alojamientos como el número de reseñas hechas en ese periodo de tiempo ha decrecido de forma notable. El primero porque muchos hoteles no pudieron soportar las pérdidas generadas y

directamente tuvieron que cerrar, y el segundo por el confinamiento y las restricciones sanitarias.

En consecuencia, muchos alojamientos turísticos se han visto en la necesidad de adaptarse a la medidas sanitarias establecidas por el Gobierno y transformar sus políticas empresariales, así como adaptar sus instalaciones y servicios para garantizar la seguridad de sus clientes y trabajadores, resultando en la generación de una actitud positiva y de confianza por parte del turista.

9. CONCLUSIONES

El objetivo principal de este estudio radicaba en la creación de un sistema que permitiera a los gestores de los alojamientos turísticos sacar conclusiones acerca de su modelo de negocio y de cómo se estaba llevando a cabo la gestión de la organización mediante un análisis de sentimientos de las valoraciones de sus clientes a nivel de aspecto, pudiendo hacerlo para un gran número de opiniones y en diferentes idiomas. Analizar aquellas características del establecimiento que se valoran de forma más positiva y aquellas valoradas de forma negativa, para, con la información obtenida, poder tomar decisiones de mejora.

Tras los análisis y resultados obtenidos, podemos sacar como conclusión principal la satisfacción general de los turistas con los alojamientos de las islas, pues la mayoría de las reseñas estudiadas presentaban un sentimiento positivo acerca de la estancia. Además, los dos aspectos valorados de forma más positiva eran las instalaciones y las vistas, seguidas del personal y la ubicación, características imprescindibles para la industria del turismo, teniendo en cuenta que pertenece al sector servicios.

Por otro lado, encontramos que los aspectos peor valorados estaban compuestos por la limpieza, el servicio y la recepción. El primero de ellos, tal y como vimos durante el desarrollo del trabajo, se considera una de las cualidades que más valoran los turistas a la hora de escoger entre alojamiento u otro. Es por ello, que se recomienda a los gerentes de los hoteles que presten mucha atención a este atributo si quieren incrementar su demanda y, con ello, las ventas e ingresos de su establecimiento.

En cuanto a la librería utilizada para la investigación y los resultados obtenidos, se puede concluir que, en general, es una buena herramienta de clasificación de sentimientos a nivel de aspecto pues, en la mayoría de los casos tenía un porcentaje alto de acierto. No obstante, existían errores en algunas reseñas, hecho esperable ya que todo sistema real de *sentiment analysis* posee un índice de desempeño inferior al 100%, como ocurre con toda solución de análisis de esta naturaleza. A pesar de ello, los errores tienen una influencia prácticamente nula en el análisis de explotación, como se corresponde con todo cuadro de mandos de ayuda a la toma de decisiones.

9.1 Principales aportaciones

a) Un beneficio importante para los alojamientos, las cadenas hoteleras, la patronal y la Comunidad Autónoma como armonizadora
b) Una aportación de tecnología de última generación al principal sector económico de las islas
c) Un producto de proyección internacional y escalable a múltiples niveles, desde la planificación hasta la gestión

Los principales beneficiarios del modelo creado son los propios alojamientos turísticos, pues tienen a su disposición una gran fuente de información lista para ser analizada y tomar decisiones de mejora en la gestión y futuras estrategias del negocio.

En primer lugar, para conocer a sus clientes y los principales aspectos que tienen en cuenta a la hora de alojarse en un hotel. Como se ha demostrado, las valoraciones positivas de los turistas suponen un incremento en consumidores potenciales y, con ello, un aumento no sólo en los ingresos del negocio, sino en su reputación y marca. Por esta razón, los gerentes de los diferentes alojamientos deberán hacer un buen uso del sistema diseñado para intentar mejorar y detectar fallos en las instalaciones o servicios.

Los turistas son el motor de la industria del turismo, la principal fuente de sustento y beneficios. Por este motivo, la satisfacción de estos constituye uno de los factores primordiales en el crecimiento y rendimiento del sector.

9.2 Trabajos futuros

En este trabajo se han analizado los casos de Gran Canaria al completo y un subconjunto de los alojamientos de Tenerife debido a limitaciones en los costes y recursos utilizados por la librería

del análisis ABSA. Es por ello, que se propone como principal trabajo futuro mejorar la escalabilidad del modelo en función de los datos recolectados en la etapa del *web scraping*.

La idea sería construir un servicio apto para múltiples clientes de diferentes partes del mundo, en el que exista la posibilidad de realizar un estudio comparativo entre alojamientos con características similares, lo que se conoce como “CompSet”.

Incluso, se podría utilizar este sistema como modelo de negocio de una empresa, donde se mitigarían los problemas de escalabilidad del modelo gracias a la adquisición, donde sea necesario, de las correspondientes licencias tanto en los procesos *web scraping* y de traducción, como en el análisis de sentimientos. El propósito de este negocio sería hacer uso de esta herramienta empresarial para proporcionar un servicio de análisis a los diferentes alojamientos turísticos.

Por otro lado, y teniendo en cuenta los resultados conseguidos a lo largo del estudio, se podrían comparar con las reseñas mostradas en otras plataformas de viajes para confirmar el correcto funcionamiento del modelo.

Finalmente, si tenemos en consideración la librería utilizada, se podrían mejorar y extender las técnicas de análisis para permitir la entrada de textos de gran tamaño, solucionando el problema de la limitación de tokens, así como perfeccionar la identificación y polaridad de aspectos. Asimismo, como futuro avance, incluiríamos la posibilidad de identificar los aspectos de forma automatizada, en lugar de ser el usuario quien los introduzca manualmente.

BIBLIOGRAFÍA

Adarsh, S., y Sreereshma, S. (2021). A study on the role of social media in tourism marketing with special reference to Kerala state. *Lux Montis*, *9*, 47-56.

Aguilar Ibáñez, A. (2017). *Desarrollo de Modelos de Deep Learning para comprensión de textos usando técnicas NLP*. Universidad de Zaragoza.

Aichner, T., y Jacob, F. (2015). Measuring the Degree of Corporate Social Media Use. *International Journal of Market Research*, *57*(2). https://doi.org/10.2501/IJMR-2015-018

Alias, G., y Cassanelli, R. (2019). *NLP aplicado a análisis de texto*. Universidad Nacional de Mar de Plata.

Andra. (2021, febrero 15). *5 Best Data Analysis Programming Languages (2021 Guide)*. https://dataresident.com/data-analysis-programming-languages/

Ankit, U. (2020, abril 24). *Transformer Neural Network: Step-By-Step Breakdown of the Beast.* https://towardsdatascience.com/transformer-neural-network-step-by-step-breakdown-of-the-beast-b3e096dc857f

BBC News Mundo. (2019, octubre 29). *Google: cómo funciona BERT, la mayor actualización del algoritmo del motor de búsqueda más usado en el mundo*. https://www.bbc.com/mundo/noticias-50223408

Bhatnagar, P. (2018, marzo 20). *Why should a hotel's online reputation matter for every hotel owner?* https://www.hotelogix.com/blog/2018/03/20/how-hotel-online-reputation-

impacts-its-revenue/

Bravo, A. (2018, abril 20). *La importancia de las redes sociales en las empresas y su gestión*. https://www.diariodemallorca.es/economia/foro-negocios-businessdm/2018/04/20/importancia-redes-sociales-empresas-gestion-3197616.html

Chan, J. (2017, agosto 6). *4 Effective Ways to Leverage Online Reviews for Your Business*. https://www.socialmediatoday.com/news/4-effective-ways-to-leverage-online-reviews-for-your-business/503002/

Chandni, Chandra, N., Gupta, S., y Pahade, R. (2015). Sentiment Analysis and its Challenges. *International Journal of Engineering Research y Technology (IJERT)*, *4*(03), 968-970.

Chaves, M. S., Gomes, R., y Pedron, C. (2012). Analysing reviews in the Web 2.0: Small and medium hotels in Portugal. *Tourism Management*, *33*(5). https://doi.org/10.1016/j.tourman.2011.11.007

Conecta Software. (s. f.). *TripAdvisor - La mayor plataforma de opiniones, precios y reservas*. https://conectasoftware.com/apps/tripadvisor/

D'Andrea, A., Ferri, F., Grifoni, P., y Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications*, *125*(3), 26-33. https://doi.org/10.5120/ijca2015905866

Dang, N. C., Moreno-García, M. N., y De la Prieta, F. (2020). Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*, *9*(3). https://doi.org/10.3390/electronics9030483

Data, R. (s. f.). *Python math Module*. https://www.w3schools.com/python/module_math.asp

Devlin, J., Chang, M.-W., Lee, K., y Toutanova, K. (2019, junio). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*. https://doi.org/10.18653/v1/N19-1423

Dialani, P. (2020, octubre 4). *NLP V/S NLU V/S NLG*. https://www.analyticsinsight.net/nlp-vs-nlu-vs-nlg/

Do, H. H., Prasad, P., Maag, A., y Alsadoon, A. (2019). Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. *Expert Systems with Applications*, *118*. https://doi.org/10.1016/j.eswa.2018.10.003

eTurboNews. (2020, septiembre 30). *The Importance of Online Reviews in Promoting Global Tourism*. https://eturbonews.com/808869/the-importance-of-online-reviews-in-promoting-global-tourism/

Europa Press. (2020, febrero 3). *Canarias cierra el año 2019 con 13,1 millones de turistas extranjeros, un 4,4% menos, y el gasto cae un 1,4%*. https://www.europapress.es/islas-canarias/noticia-canarias-cierra-ano-2019-131-millones-turistas-extranjeros-44-menos-gasto-cae-14-20200203121759.html

FRONTUR CANARIAS. (2020). *Turistas en Gran Canaria 2010-2020*.

González García, C., Núñez-Valdez, E., García-Díaz, V., Pelayo G-Bustelo, C., y Cueva-Lovelle, J. M. (2019). A Review of Artificial Intelligence in the Internet of Things. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(4), 9. https://doi.org/10.9781/ijimai.2018.03.004

Gonzalo, F. (2014a, febrero 14). *How TripAdvisor Impacts Travel Decision-Making [INFOGRAPHIC]*. https://www.socialmediatoday.com/news/how-tripadvisor-impacts-travel-decision-making-infographic/460467/

Gonzalo, F. (2014b, marzo 30). *Why Review Sites are a Must for Travel Marketers*. https://www.socialmediatoday.com/news/why-review-sites-are-a-must-for-travel-marketers/459089/

Google Cloud. (s. f.). *Cloud Translation*. https://cloud.google.com/translate/?hl=es

Google Colaboratory. (2017). *Te damos la bienvenida a Colaboratory*. https://colab.research.google.com/notebooks/intro.ipynb

Han, S. (2020). *Googletrans Documentation Release 3.0.0*.

Henríquez, C., Pla, F., Hurtado, L.-F., y Guzmán, J. (2017). Análisis de sentimientos a nivel de aspecto usando ontologías y aprendizaje automático. *Procesamiento del Lenguaje Natural*, *59*, 49-56.

Hola Islas Canarias. (2021). *Bono Turístico Islas Canarias*. https://www.holaislascanarias.com/bono-turistico/

IBM. (2020). *Conceptos básicos de ayuda de CRISP-DM*. https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=dm-crisp-help-overview

IBM Cloud Education. (2020, mayo 27). *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?* https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

Icoz, O., Kutuk, A., y Icoz, O. (2018). *Ps418_12*. *16*, 1051-1066.

ILTADMIN. (2013, marzo 20). *Impact of Social Media on Travel Decision Making*. https://www.indianluxurytrains.com/blog/impact-of-social-media-on-travel-industry/

Indurkhya, N., y Damerau, F. J. (2010). *Handbook of Natural Language Processing* (2nd Edition). Chapman and Hall/CRC. https://doi.org/10.1201/9781420085938

JetBrains. (2020). *Python Developers Survey 2020 Results*. https://www.jetbrains.com/lp/python-developers-survey-2020/

Jupyter. (2018). *Project Jupyter*. https://jupyter.org/

Kasaraneni, C. K. (2020, junio 24). *Understanding NLP Pipeline*. https://medium.com/analytics-vidhya/understanding-nlp-pipeline-9af8cba78a56

Katrekar, A. (2019). *An Introduction to Sentiment Analysis*.

Kaur, J. (2021, enero 15). *What are the Differences Between NLP, NLU, and NLG?* https://www.xenonstack.com/blog/difference-between-nlp-nlu-nlg

Kovács, Z., Vida, G., Elekes, Á., y Kovalcsik, T. (2021). Combining Social Media and Mobile Positioning Data in the Analysis of Tourist Flows: A Case Study from Szeged, Hungary. *Sustainability*, *13*(5). https://doi.org/10.3390/su13052926

León Martínez, J. (2019). *Identificación de depresión mediante el análisis de sentimientos*. Universidad de Extremadura.

Liu, B. (2015). *Sentiment analysis: mining opinions, sentiments, and emotions / Bing Liu.* https://search.ebscohost.com/login.aspx?direct=trueydb=cat07429ayAN=ulpgc.757235y site=eds-live

Lou, C., y Yuan, S. (2019). Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media. *Journal of Interactive Advertising*, *19*(1). https://doi.org/10.1080/15252019.2018.1533501

Machado Chaviano, E. L., y Hernández Aro, Y. (2008). Del turismo contemplativo al turismo activo. *El Periplo Sustentable*, *15*. https://doi.org/10.21854/eps.v0i15.937

Mahler, T., Cheung, W., Elsner, M., King, D., de Marneffe, M.-C., Shain, C., Stevens-Guille, S., y White, M. (2017, septiembre). Breaking NLP: Using Morphosyntax, Semantics, Pragmatics and World Knowledge to Fool Sentiment Analysis Systems. *Proceedings of the First Workshop on Building Linguistically Generalizable NLP Systems*. https://doi.org/10.18653/v1/W17-5405

Mallamma, V. R., y Hanumanthappa, M. (2014). Semantical and Syntactical Analysis of NLP. *International Journal of Computer Science and Information Technologies*, *5*(3), 3236-3238.

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez Orallo, J., Kull, M., Lachiche, N., Ramirez Quintana, M. J., y Flach, P. A. (2019). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*. https://doi.org/10.1109/TKDE.2019.2962680

Matilla López, M., Ríos Martín, M. Á., y Ortega Fraile, F. J. (2016). Catalogación de los aspectos más relevantes según los comentarios de Tripadvisor al elegir un hotel en Sevilla. *El turismo y la experiencia del cliente: IX jornadas de investigación en turismo*, 91-123.

McKinney, W. (2008). *pandas - Python Data Analysis Library*. https://pandas.pydata.org/

Microsoft Docs. (2021, abril 2). *Uso de Conclusiones de IA en Power BI Desktop*. https://docs.microsoft.com/es-es/power-bi/transform-model/desktop-ai-insights

Miguéns, J., Baggio, R., y Costa, C. (2008). Social media and tourism destinations: TripAdvisor case study. *Advances in tourism research*, *26*(28), 1-6.

Mkono, M., y Tribe, J. (2017). Beyond Reviewing: Uncovering the Multiple Roles of Tourism Social Media Users. *Journal of Travel Research*, *56*(3). https://doi.org/10.1177/0047287516636236

MonkeyLearn. (s. f.). *Sentiment Analysis: A Definitive Guide*. https://monkeylearn.com/sentiment-analysis/#how-does-sentiment-analysis-work

Naumenko, V. (s. f.). *Top Data Science Programming Languages*. https://jelvix.com/blog/top-data-science-programming-languages

Naushan, H. (2020, octubre 1). *Sentiment Analysis of Social Media with Python | Towards Data Science*. https://towardsdatascience.com/sentiment-analysis-of-social-media-with-python-45268dc8f23f

Negrut, V. (2018). POWER BI: EFFECTIVE DATA AGGREGATION. *Quaestus*, *13*, 146-152. https://powerbi.microsoft.com

Panico, C. (2018). *La eficacia del análisis de sentimientos para la empresa: el caso de estudio Dell Technologies Inc.* Universidad Complutense de Madrid.

Pascual, F. (2019, marzo 8). *Guide to Aspect-Based Sentiment Analysis*. https://monkeylearn.com/blog/aspect-based-sentiment-analysis/

Patel, S. (2020a, julio 17). *NLP Guide 101: NLP Evolution — Past, Present and Future.…*. https://suneelpatel-in.medium.com/nlp-guide-101-nlp-evolution-past-present-and-future-fcc573629da3

Patel, S. (2020b, agosto 2). *NLP Pipeline: Building an NLP Pipeline, Step-by-Step*. https://suneelpatel-in.medium.com/nlp-pipeline-building-an-nlp-pipeline-step-by-step-7f0576e11d08

Pauli, P. A. (2019). *Análisis de sentimiento: Comparación de algoritmos predictivos y métodos utilizando un lexicon español*. Instituto Tecnológico de Buenos Aires - ITBA.

Paús, F., y Macchia, L. (2014). MARKETING VIRAL EN MEDIOS SOCIALES: ¿QUÉ CONTENIDO ES MÁS CONTAGIOSO Y POR QUÉ? *Ciencias Administrativas*, *4*, 67-82. https://www.redalyc.org/articulo.oa?id=511651380007

Poria, S., Chaturvedi, I., Cambria, E., y Bisio, F. (2016, julio). Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. *2016 International Joint Conference on Neural Networks (IJCNN)*. https://doi.org/10.1109/IJCNN.2016.7727784

Preethi, G., Krishna, P. V., Obaidat, M. S., Saritha, V., y Yenduri, S. (2017, julio). Application of Deep Learning to Sentiment Analysis for recommender system on cloud. *2017 International Conference on Computer, Information and Telecommunication Systems (CITS)*. https://doi.org/10.1109/CITS.2017.8035341

Promotur Turismo de Canarias, y Gobierno de Canarias. (2018). *Plan de marketing estratégico 2018 - 2022*. https://turismodeislascanarias.com/sites/default/files/plan_de_marketing_estrategico_2018-2022_0.pdf

Purkayastha, S. (2019, junio 19). *Top 10 Best Translation APIs in 2021*. https://blog.api.rakuten.net/top-10-best-translation-apis-google-translate-microsoft-

translator-and-others/#Top_10_Best_APIs_for_Translation

Python Docs. (s. f.-a). *math — Mathematical functions*. https://docs.python.org/3/library/math.html

Python Docs. (s. f.-b). *random — Generate pseudo-random numbers*. https://docs.python.org/3/library/random.html

Python Docs. (s. f.-c). *time — Time access and conversions*. https://docs.python.org/es/3/library/time.html

Quonext. (s. f.). *Microsoft-Power-BI-folleto-Quonext*. https://www.quonext.com/descargables/Microsoft-Power-BI-folleto-Quonext.pdf

Reitz, A. K. (2013). *Requests: HTTP para Humanos — documentación de Requests - 1.1.0*. https://docs.python-requests.org/es/latest/

Repustate. (s. f.). *Sentiment Analysis Applications In Business*. https://www.repustate.com/sentiment-analysis-solutions/

Richardson, L. (2020). *Beautiful Soup Documentation*. https://www.crummy.com/software/BeautifulSoup/bs4/doc/

Rico Shmidt, E. (2013). *json — Notación de objetos JavaScript — El módulo Python 3 de la semana*. https://rico-schmidt.name/pymotw-3/json/

Rodrigo García, G. (2014). *Análisis de dos cadenas hoteleras. Un estudio comparativo de Nh Hoteles y Melia Hotels International* [Universidad de Valladolid]. http://uvadoc.uva.es/handle/10324/6005

Rolczynski, R. (2020, diciembre 14). *aspect-based-sentiment-analysis · PyPI*. https://pypi.org/project/aspect-based-sentiment-analysis/

Rolczyński, R. (s. f.). *Do You Trust in Aspect-Based Sentiment Analysis? Testing and Explaining Model Behaviors Table Of Content*.

Sakunkoo, P., y Sakunkoo, N. (2009, mayo). Analysis of Social Influence in Online Book Reviews. *Proceedings of the International AAAI Conference on Web and Social Media*.

Salvi, F., Serra Cantallops, A., y Ramón Cardona, J. (2013). Los impactos del ewom en hoteles. *Redmarka. Revista de Marketing Aplicado*, *2*(010), 3-17. https://doi.org/10.17979/redma.2013.02.010.4765

Sánchez-Amboage, E., Rodríguez -Fernández, M. M., Juanatey-Boga, Ó., y Martínez-Fernández, V. A. (2019). La comunicación de los destinos turísticos en los medios sociales: el caso de la España Verde. *Revista Espacios*, *40*(11), 11-undefined.

Sarmiento Guede, J., de Esteban Curiel, J., y Antonovica, A. (2017). La comunicación viral a través de los medios sociales: análisis de sus antecedentes. *Revista Latina de Comunicación Social*, *72*, 69-86. https://doi.org/10.4185/RLCS-2017-1154

Satapathy, R., Cambria, E., y Hussain, A. (2017). *Sentiment Analysis in the Bio-Medical Domain* (Vol. 7). Springer International Publishing. https://doi.org/10.1007/978-3-319-68468-0

Schouten, K., y Frasincar, F. (2016). Survey on Aspect-Level Sentiment Analysis. *IEEE Transactions on Knowledge and Data Engineering*, *28*(3). https://doi.org/10.1109/TKDE.2015.2485209

Sellés Revert, R. (2016). *El uso de las redes sociales en el ámbito empresarial: análisis de los determinantes de su adopción, intensidad de uso e influencia*. Universitat de València.

Shahnawaz, y Astya, P. (2017, mayo). Sentiment analysis: Approaches and open issues. *2017 International Conference on Computing, Communication and Automation (ICCCA)*. https://doi.org/10.1109/CCAA.2017.8229791

Sharma, A. (2020, marzo 23). *Use Google Colab for Deep Learning and Machine Learning Models*. https://www.analyticsvidhya.com/blog/2020/03/google-colab-machine-learning-deep-learning/

Silveira, K. K. B., Pereira, L. A., y Limberger, P. F. (2021). Social media standardization assessment managed by the Ministryof Tourism. *Marketing y Tourism Review*, *6*(1). https://doi.org/10.29149/mtr.v6i1.6352

Stringam, B. B., Gerdes, J., y Vanleeuwen, D. M. (2010). Assessing the Importance and Relationships of Ratings on User-Generated Traveler Reviews. *Journal of Quality Assurance in Hospitality y Tourism*, *11*(2). https://doi.org/10.1080/1528008X.2010.482000

Suau Jiménez, F. (2012). El turista 2.0 como receptor de la promoción turística: estrategias lingüísticas e importancia de su estudio. *PASOS. Revista de Turismo y Patrimonio Cultural*, *10*(4). https://doi.org/10.25145/j.pasos.2012.10.060

Sugiharti, D., Chaiechi, T., y Pryce, J. (2021). *The Role of Visitor's Resilience in Understanding Tourism Resilience: a Conceptual Framework*. https://www.researchgate.net/publication/350088703

Sütçü, C. S., y Aytekin, Ç. (2019). An Example of Pragmatic Analysis in Natural Language Processing: Sentimental Analysis of Movie Reviews. *Communication and Technology Congress – CTC 2019*, 61-74.

Taecharungroj, V., y Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, *75*. https://doi.org/10.1016/j.tourman.2019.06.020

Tarasov, D. S. (2015). *Natural Language Generation, Paraphrasing and Summarization of User Reviews with Recurrent Neural Networks*.

Tatan, V. (s. f.). *Intro to Google Colab for Data Analytics*. https://towardsdatascience.com/intro-to-google-colab-for-data-analytics-da5e3a37af8a

Tavva, R. (2021, marzo). *Natural Language Processing Pipelines, Explained*. https://www.kdnuggets.com/2021/03/natural-language-processing-pipelines-explained.html

tourinews. (2020, febrero 13). *Islas Canarias, a la caza de turistas italianos en la BIT de Milán*. https://www.tourinews.es/eventos/islas-canarias-turismo-italiano-bit-milan_4459000_102.html

Tripadvisor Insights. (2014, febrero 11). *24 insights to shape your Tripadvisor strategy*. https://www.tripadvisor.com/TripAdvisorInsights/w710

Tuominen, P. (2011). *The influence of TripAdvisor consumer-generated travel reviews on hotel performance*.

Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A. N., Gouws, S., Jones, L., Kaiser, Ł., Kalchbrenner, N., Parmar, N., Sepassi, R., Shazeer, N., y Uszkoreit, J. (2018). *Tensor2Tensor for Neural Machine Translation*. http://arxiv.org/abs/1803.07416

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. http://arxiv.org/abs/1706.03762

Wang, D., Zhu, S., y Li, T. (2013). SumView: A Web-based engine for summarizing product reviews and customer opinions. *Expert Systems with Applications*, *40*(1). https://doi.org/10.1016/j.eswa.2012.05.070

Xie, K. L., Zhang, Z., Zhang, Z., Singh, A., y Lee, S. K. (2016). Effects of managerial response on consumer eWOM and hotel performance. *International Journal of Contemporary Hospitality Management*, *28*(9). https://doi.org/10.1108/IJCHM-06-2015-0290

Ye, Q., Law, R., y Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1). https://doi.org/https://doi.org/10.1016/j.ijhm.2008.06.011

Yogesh, F., y Yesha, M. (2014). Effect of Social Media on Purchase Decision. *Pacific Business Review International*, *6*(11), 45-51.

Yuzdepski, Z. (s. f.). *Needs-Based Selling: The 5 Phases of the Modern Customer Journey*. https://www.vendasta.com/blog/following-modern-customer-journey/

ANEXO

- Visualización de resultados en Power BI: Visualización TFG - Power BI