



**GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE
EMPRESAS**

CURSO ACADÉMICO 2018 - 2019

TRABAJO FIN DE GRADO

**Aplicación de técnicas de *Opinion Mining* para el
análisis de la reputación online en el sector
hotelero**

**Application of *Opinion Mining* techniques for the
analysis of online reputation in the hotel sector**

AUTORA: VALERIA COBO ROCHA

DIRECTORA: RAQUEL GÓMEZ LÓPEZ

Santander, 28 de Junio de 2019

TABLA DE CONTENIDO

Resumen	4
Abstract	4
1. MARCO GENERAL DEL TRABAJO	5
1.1. Introducción.....	5
1.2. Justificación.....	5
1.3. Objetivos del trabajo	6
1.3.1. <i>Objetivo general</i>	6
1.3.2. <i>Objetivos específicos</i>	6
2. MARCO TEÓRICO	6
2.1. El sector turístico en España	6
2.2. Las TIC en el sector turístico	7
2.3. Consumidor 2.0: el perfil del consumidor en un contexto digital	8
2.4. Reputación online	10
2.4.1. <i>Herramientas para la adecuada gestión de la reputación online</i>	10
2.4.2. <i>El papel del community manager</i>	11
3. METODOLOGÍA	12
3.1. Análisis automatizado de opiniones en redes sociales	12
3.2. Opinion Mining (OM).....	12
3.3. Herramientas utilizadas	14
4. DESARROLLO EMPÍRICO	14
4.1. Ejemplo de monitorización de menciones en Internet: caso Hotel Real de Santander....	15
4.2. Análisis de sentimiento de los hoteles de lujo de la ciudad de Santander.....	19
4.2.1. <i>Extracción de opiniones</i>	19
4.2.2. <i>Preprocesamiento</i>	20
4.2.3. <i>Identificación de sentimientos</i>	21
4.2.4. <i>Clasificación de opinión</i>	23
4.3. Respuestas de los hoteles ante los comentarios de los huéspedes	29
5. CONCLUSIONES	30
6. LIMITACIONES Y LÍNEAS FUTURAS DE TRABAJO	31
REFERENCIAS	32

ÍNDICE DE FIGURAS

FIGURA 2.1. PERNOCTACIONES SEGÚN TIPO DE ALOJAMIENTO (MILLONES). FUENTE: ESPAÑA EN CIFRAS 2018. INSTITUTO NACIONAL DE ESTADÍSTICA (INE, 2018B)	7
FIGURA 2.2. FUENTES DE INFORMACIÓN PARA LA COMPRA DE PRODUCTOS TURÍSTICOS POR PARTE DE LOS CONSUMIDORES. FUENTE: (CETELM, 2018).	9
FIGURA 3.1. EJEMPLOS DE CLASIFICACIÓN DE SENTIMIENTO DE OPINIONES CON LA HERRAMIENTA MR.TUIT.....	13
FIGURA 3.2. FASES DE TODO PROCESO DE OPINION MINING.	14
FIGURA 4.1. HOTELES DE LA CIUDAD DE SANTANDER OBJETO DE ANÁLISIS.	14
FIGURA 4.2. VALORACIONES DE CLIENTES DEL HOTEL REAL DE SANTANDER A TRAVÉS DE BOOKING.	15
FIGURA 4.3. OPINIONES A TRAVÉS DE LA PLATAFORMA TRIPADVISOR.....	16
FIGURA 4.4. HERRAMIENTA BRAND24 PARA LA MONITORIZACIÓN DE LA PRESENCIA WEB.	17
FIGURA 4.5. INDICADORES SOBRE LA PRESENCIA WEB DEL HOTEL REAL DE SANTANDER A TRAVÉS DE LA PLATAFORMA BRAND24.	17
FIGURA 4.6. RESUMEN DE LAS FUENTES DE INFORMACIÓN CON MENCIONES AL HOTEL REAL DE SANTANDER.	18
FIGURA 4.7. EJEMPLO DE MENCIÓN CALIFICADA COMO “POSITIVA” POR BRAND24.	18

FIGURA 4.8. EJEMPLO DE MENCIÓN CALIFICADA COMO “NEGATIVA” POR BRAND24.	18
FIGURA 4.9. EJEMPLO DE MENCIÓN CALIFICADA COMO “POSITIVA” POR BRAND24 PROVENIENTE DE PRENSA DIGITAL.	19
FIGURA 4.10. PREPROCESAMIENTO DE LAS OPINIONES.	21
FIGURA 4.11. EJEMPLO DE CÁLCULO DE LA PUNTUACIÓN DE UNA OPINIÓN.	22
FIGURA 4.12. CORRELACIÓN POSITIVA ENTRE PUNTUACIONES OM Y VALORACIONES DIRECTAS EN TRIPADVISOR.	24
FIGURA 4.13. PUNTUACIONES MEDIAS OBTENIDAS POR LOS HOTELES DE LUJO DE SANTANDER.	25
FIGURA 4.14. PROPORCIÓN DE OPINIONES CON SENTIMIENTO POSITIVO/NEGATIVO EN LOS HOTELES ANALIZADOS.	25
FIGURA 4.15. PALABRAS POSITIVAS MÁS FRECUENTES EN LAS OPINIONES DE LOS HOTELES.	26
FIGURA 4.16. PALABRAS NEGATIVAS MÁS FRECUENTES EN LAS OPINIONES DE LOS HOTELES.	27
FIGURA 4.17. PALABRAS MÁS FRECUENTES EN LOS COMENTARIOS POSITIVOS SOBRE EL HOTEL REAL.	28
FIGURA 4.18. PALABRAS MÁS FRECUENTES EN LOS COMENTARIOS NEGATIVOS SOBRE EL HOTEL REAL.	29

ÍNDICE DE TABLAS

TABLA 2.1. TIPOS DE ALOJAMIENTOS TURÍSTICOS ABIERTOS. FUENTE: ESPAÑA EN CIFRAS 2018. INSTITUTO NACIONAL DE ESTADÍSTICA (INE, 2018B).	7
TABLA 2.2. IMPLANTACIÓN DE LAS TIC EN EL SECTOR HOTELERO. FUENTE: (ONTSI, 2018)	8
TABLA 4.1. RESUMEN DE LAS OPINIONES EXTRAÍDAS SOBRE HOTELES 4 Y 5* DE SANTANDER.	20
TABLA 4.2. EJEMPLOS DE PALABRAS DEL LÉXICO AFINN.	21
TABLA 4.3. COMPARACIÓN ENTRE LAS VALORACIONES DIRECTAS DE LOS HUÉSPEDES Y LAS PUNTUACIONES OBTENIDAS CON TÉCNICAS OPINION MINING (OM).	23
TABLA 4.4. TASAS DE RESPUESTA DE LOS HOTELES DE SANTANDER A LOS COMENTARIOS DE SUS USUARIOS.	30

RESUMEN

La reputación online en el sector hotelero tiene unas grandes implicaciones desde el punto de vista de los resultados empresariales. Los clientes de hoteles son cada vez más activos en las redes sociales y se informan por esos medios antes de realizar reservas y contrataciones de paquetes turísticos. Por ello, los hoteles deberían prestar atención a las percepciones y comentarios de sus clientes en Internet. Teniendo en cuenta los grandes volúmenes de información que se generan cada día en blogs, foros, redes sociales, y servicios online específicos del sector turístico, resulta esencial contar con herramientas que permitan automatizar su gestión. Por todo lo anterior, el objetivo de este trabajo es mostrar a los gestores de hoteles cómo el uso de herramientas informáticas y computacionales les puede ayudar en la gestión eficaz de los comentarios y opiniones que dejan sus huéspedes en Internet. Para la consecución de este objetivo se ha realizado un análisis de opiniones de huéspedes de hoteles de lujo de la ciudad de Santander, a través de técnicas de procesamiento de lenguaje natural y de análisis de sentimiento.

ABSTRACT

From the point of view of business results, online reputation in the hotel sector has great implications. Hotel clients are increasingly active in social media platforms and they search information using these sources before making a booking or hiring holiday packages. Therefore, hotels should pay attention to the perceptions and comments of their consumers on the Internet. Taking into account the large volumes of information that are generated every day in blogs, forums, social networks, and online services specific to the tourism sector, it is essential to have tools to automate the information management. Due to the aforesaid, the aim of this project is to show hotel managers how to use computational tools in order to help the effective management of comments and opinions left by their guests on the Internet. To achieve this goal, an analysis of the opinions of guests of luxury hotels in the city of Santander was carried out using natural language processing techniques and sentiment analysis.

1. MARCO GENERAL DEL TRABAJO

1.1. INTRODUCCIÓN

La reputación corporativa puede verse como el conjunto de percepciones sobre la empresa de los diferentes grupos de interés con los que se relaciona (clientes, proveedores, empleados, accionistas, etc.). Esa reputación es el resultado del comportamiento desarrollado por la empresa a lo largo del tiempo y su capacidad de creación de valor para esos grupos, tal como establece el *Corporate Excellence* y el *Reputation Institute*.

En las últimas décadas, como consecuencia de la transformación digital y la normalización de las redes sociales, la reputación corporativa ha sufrido grandes cambios. En un contexto digital, la reputación online se forma a partir de un conjunto de elementos que representan el prestigio de una organización o persona en la red. A diferencia de una marca, que puede ser creada y controlada a través de la publicidad, la reputación online escapa al control de la organización. Esa reputación es creada por los diferentes agentes implicados (clientes, trabajadores, proveedores,...) cuando conversan y aportan sus opiniones a través de medios digitales.

El sector turístico es uno de los de mayor desarrollo y madurez en el comercio electrónico y las comunidades virtuales son un factor estratégico de mejora continua de las empresas turísticas. Más allá de factores como el precio, los comentarios online de los usuarios se han convertido en una fuente predominante de influencia a la hora de tomar una decisión de compra (Pabel y Prideaux, 2016). Por ello, en el sector hotelero las nuevas formas de gestionar las relaciones con los clientes deben preocuparse por gestionar de manera eficiente los comentarios y opiniones de los clientes en redes sociales, foros de Internet, blogs, etc.

1.2. JUSTIFICACIÓN

Las empresas turísticas y en particular los hoteles, deben ser capaces de realizar un análisis adecuado de la imagen que sus clientes crean en el contexto online a través de sus comentarios. Ante las enormes cantidades de datos textuales disponibles en Internet y el uso masivo de los medios sociales de comunicación online, los hoteles necesitan recurrir a herramientas y metodologías que les faciliten la automatización de la recepción y análisis de las opiniones de sus clientes.

El presente Trabajo Fin de Grado pretende presentar distintos instrumentos que pueden ser de interés para la gestión de la reputación online y especialmente en el análisis del sentimiento que los clientes manifiestan a través de sus comentarios. Gracias a estos análisis, las empresas puede orientar sus estrategias hacia mejorar aquellos aspectos que generan mayores sentimientos negativos o de insatisfacción entre sus clientes. Por otro lado, también es interesante identificar aquellos sentimientos positivos y los factores que influyen en esas percepciones positivas o los perfiles de usuario que mejor valoran los servicios del hotel, con objeto de identificar posibles situaciones de ventaja competitiva o nichos de mercado.

1.3. OBJETIVOS DEL TRABAJO

1.3.1. Objetivo general

Este trabajo pretende servir de guía a los gestores de hoteles a la hora de gestionar de manera eficaz los comentarios y opiniones que los huéspedes dejan en Internet y analizar los posibles efectos de una mala reputación online en los resultados empresariales.

1.3.2. Objetivos específicos

El objetivo general puede desagregarse en una serie de objetivos específicos:

- Presentar herramientas tecnológicas que permitan automatizar la recepción y búsqueda de opiniones y comentarios sobre hoteles y analizar la reputación online de los mismos.
- Utilizar técnicas de Procesamiento del Lenguaje Natural (PLN) para poder clasificar las opiniones en base al sentimiento que transmiten.
- Generar informes con técnicas de *Visual Analytics* para comprender cómo los comentarios están contribuyendo a la generación de la reputación online del hotel.
- Realizar comparativas entre hoteles que se orientan hacia un mismo segmento de mercado.
- Ofrecer/Realizar recomendaciones a los responsables de los hoteles para la adecuada gestión de su presencia digital.

2. MARCO TEÓRICO

2.1. EL SECTOR TURÍSTICO EN ESPAÑA

No cabe duda de que el sector turístico en España es uno de los sectores con mayor importancia desde el punto de vista económico, social y cultural.

El turismo es el motor de la economía española y su contribución al PIB supone el 11,91% según el último informe elaborado por la Cátedra de Desarrollo Turístico Sostenible de la Universidad Nebrija y la Mesa del Turismo (Figuerola, 2018). En dicho informe se pone de manifiesto también un aumento del 11,43% de las pernoctaciones hoteleras en el periodo 2014-2018 y una aportación del sector turístico a la balanza fiscal de 24.000 millones de euros.

En términos de empleo, el sector es capaz de generar 2,6 millones de puestos de trabajo, lo que representa el 12,8% del empleo total, según los últimos datos disponibles de la Cuenta Satélite del Turismo en España del INE (2018). Por otro lado, en relación a los últimos datos aportados por el INE (2018b), correspondientes al año 2017, se produjeron en nuestro país más de 471 millones de pernoctaciones, un 72% correspondientes a estancias en hoteles, y con una estancia media de 3,7 pernoctaciones por viajero. Además, tal como se observa en la Figura 2.1, la tendencia a las pernoctaciones hoteleras es creciente. Otro dato llamativo del informe es el hecho que un 23,1% de las pernoctaciones en hoteles y establecimientos similares de la Unión Europea se realizan en España.

La oferta hotelera española, tomando como referencia el informe España en Cifras 2018 (INE, 2018b), se concreta en los establecimientos que se incluyen en la Tabla

2.1. Como puede verse, la estimación de hoteles con los que cuenta nuestro país supera los 14.000.

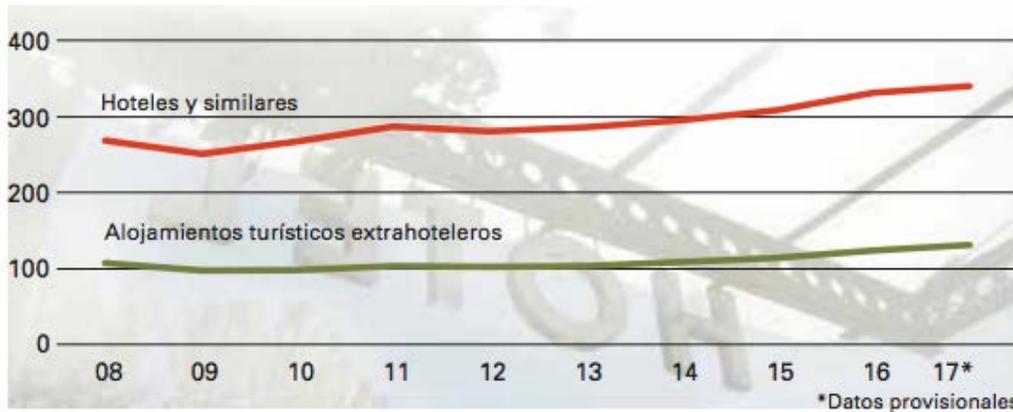


Figura 2.1. Pernoctaciones según tipo de alojamiento (millones). Fuente: España en cifras 2018. Instituto Nacional de Estadística (INE, 2018b)

Tipo de alojamiento	Establecimientos estimados (medias anuales)
Hoteles	14.659
Apartamentos turísticos	134.561
Campings	763
Alojamientos de turismo rural	16.283
Albergues	217

Tabla 2.1. Tipos de alojamientos turísticos abiertos. Fuente: España en cifras 2018. Instituto Nacional de Estadística (INE, 2018b).

2.2. LAS TIC EN EL SECTOR TURÍSTICO

En el ámbito empresarial, la digitalización ha dejado de ser un factor estratégico de competitividad para convertirse en un requisito esencial para la propia supervivencia empresarial (Fundación Telefónica, 2019). Las empresas del sector turístico no son ajenas a esta necesidad y, de hecho, ya en 2015 se ponía de manifiesto que el sector turístico se encontraba entre los cinco sectores de mayor madurez digital, sobretodo en relación al uso de tecnologías digitales para interaccionar con clientes y empleados y en la definición de estrategias de transformación y competencias de gestión (Fundación Orange, 2016; Kane *et al.*, 2015).

Las empresas hoteleras son conscientes de la importancia de tener una presencia Web y de estar presentes en las redes sociales. Según el informe e-pyme 2017 en el que se analiza la implantación de las TIC en las empresas españolas (ONTSI, 2018), en el sector de hoteles y agencias de viajes prácticamente el 100% de las empresas cuentan con conexión a Internet, el 94,6% de las PYMES y grandes empresas disponen de página web corporativa, bajando ese porcentaje a un 64,1% en el caso de las microempresas. En cuanto al uso de redes sociales, no hay tanta diferencia en función del tamaño de las empresas, un 98,3% de las PYMES y grandes empresas hacen uso de redes sociales y en el caso de las microempresas el porcentaje sería un 98,2%. La Tabla 2.2 muestra algunos datos adicionales de interés que ponen de manifiesto la alta implantación de las TIC en el sector hotelero en España.

	% PYMES y grandes empresas	% Microempresas
Disponibilidad de ordenador	100	94,1
Conexión a Internet	100	92,8
Página web corporativa	94,6	64,1
Recepción de reservas a través de la página web	85,5	44,2
Uso de herramientas CRM	49,2	14,2
Uso de redes sociales	98,3	98,2
Utilización de webs para compartir contenido multimedia	56,7	37,7
Mantenimiento de un blog empresarial	49,4	36,6
Implantación de medidas de ciberseguridad	92,9	60,4
Uso de publicidad dirigida basada en contenido web o búsquedas de usuarios en Internet	87,9	64,0
Uso de publicidad basada en rastreo de actividades anteriores de usuarios	49,5	25,4
Uso de publicidad dirigida basada en geolocalización de usuarios de Internet	47,0	25,5
Analizan datos generados por medios sociales	78,1	78,1

Tabla 2.2. Implantación de las TIC en el sector hotelero. Fuente: (ONTSI, 2018)

Según la Fundación Orange (2016), un 13,2% del volumen de negocio del comercio electrónico en España corresponde a agencias de viajes y operadores turísticos. Ocio, turismo y artículos de moda se encuentran a la cabeza de las compras por Internet en nuestro país. Según el último informe del Observatorio Cetelem de E-commerce 2018, el 73% de los usuarios de Internet declararon haber realizado alguna compra relacionada con servicios turísticos y viajes en el último año. Entre los productos comprados online destacan la reserva de hoteles (un 56% de los consumidores) y de vuelos (50%) (Cetelem, 2018).

2.3. CONSUMIDOR 2.0: EL PERFIL DEL CONSUMIDOR EN UN CONTEXTO DIGITAL

Las redes sociales, y en general la conocida como Web 2.0, han cambiado la forma en la que las personas y las empresas interactúan entre sí. Las plataformas sociales permiten a los usuarios generar contenido, compartirlo, manifestar opiniones y en general, crear verdaderas comunidades de usuarios. La capacidad de interaccionar con otros consumidores y crear contenido da a los propios consumidores un poder que hasta ese momento no tenían y obliga a las empresas a plantearse nuevas formas para atraer a los clientes y gestionar sus relaciones con ellos. Así, por ejemplo, el sector hotelero y de agencias de viajes es el que concentra la mayor proporción de compañías que invierten en publicitarse en la Web (ONTSI, 2018).

En dicho contexto digital, los clientes del sector hotelero también hacen uso de estas tecnologías para informarse de los destinos, elegir los alojamientos, realizar reservas o planificar las actividades a realizar durante sus viajes.

En primer lugar puede ser interesante estudiar a través de que tipo de página web los consumidores optan por hacer sus reservas de hoteles. Según el informe (Cetelem, 2018), el medio más utilizado es la propia página web del hotel (46% de los

consumidores), seguido de los comparadores de precios (36%) y las agencias de viajes online (33%). La facilidad para acceder a diferentes ofertas o fuentes de información a través de Internet hace que el consumidor antes de realizar su compra visite diferentes webs, acceda a comentarios de otros clientes y compare precios. De hecho, en ese mismo informe se estima que el 48% de los compradores consulta 3 o más webs antes de contratar un viaje o reservar un hotel.

El informe (Cetelem, 2018) muestra también información interesante sobre la forma en la que los usuarios se informan antes de comprar un producto turístico. La Figura 2.2 muestra información al respecto de este asunto. Como hecho destacado, se podría decir que al menos el 38% de los consumidores se basan en opiniones de usuarios de Internet, amigos o familiares para comprar un producto turístico. Probablemente en el caso específico de hoteles, este porcentaje sea mucho más alto, debido principalmente a la gran cantidad de información disponible sobre hoteles y a la existencia de portales específicos para comparar precios y opiniones.

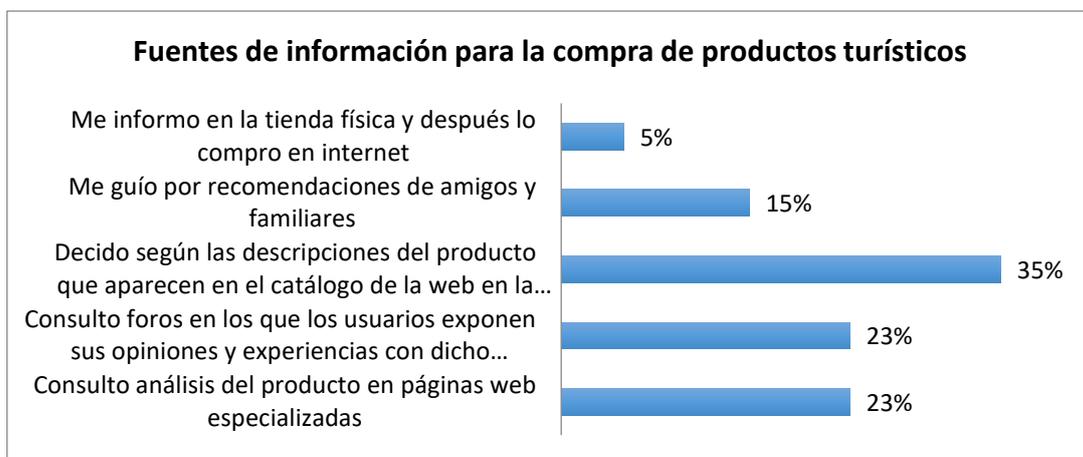


Figura 2.2. Fuentes de información para la compra de productos turísticos por parte de los consumidores. Fuente: (Cetelem, 2018).

Otros estudios también vienen a confirmar el hecho de que los consumidores muchas veces prefieren conocer opiniones reales de otros consumidores antes que los propios materiales de marketing de los hoteles (Filleri *et al.*, 2015; Herrero *et al.*, 2015).

En el último estudio de *TripAdvisor* (2018) sobre una muestra de 23.198 viajeros se pone de manifiesto que el 44% de ellos declararon que *TripAdvisor* fue una fuente de información para la elección de su alojamiento. El 21% de los viajeros también declaraba el uso de buscadores de Internet y el 13% se basaba en contenido generado por otros usuarios. Incluso, el estudio indica que uno de cada diez viajeros cambia el destino que inicialmente tenía en mente durante el proceso de búsqueda de información en Internet.

En definitiva, lo conocido como *User-Generated Content* (UGC) viene a tener una muy destacada influencia a la hora de tomar decisiones sobre la reserva de hoteles. Además el volumen del UGC generado diariamente crece de manera exponencial. Por ejemplo, *Instagram* reportó la generación de más de 95 millones de fotos y videos diariamente durante 2017; en 2018 en *TripAdvisor* se incluían 255 nuevas opiniones cada minuto; en julio de 2017 *Booking.com* contaba con 123 millones de reseñas sobre hoteles. A todo esto, debe sumarse la información compartida por usuarios a través de redes sociales de carácter más general como *Twitter*, *Facebook* o *Google+*. Toda esta información afecta en gran medida a la imagen corporativa de las entidades hoteleras.

2.4. REPUTACIÓN ONLINE

La reputación es una faceta que afecta al posicionamiento de las empresas, entendido éste como la propuesta de valor o ventaja comparativa de las mismas (Rubio *et al.*, 2017). Las consecuencias de una reputación positiva pueden ser diferentes, desde facilitar el proceso de selección por los consumidores (Walsh *et al.*, 2009), hasta atraer a mejores empleados (Fombrum *et al.*, 2000). Una buena reputación permite también tener menores gastos de marketing y mantener precios más altos. Por ejemplo, una buena reputación online puede hacer que la web de la empresa sea considerada una página de autoridad en su sector y debido al funcionamiento de los buscadores de Internet recibirá un mayor volumen de tráfico orgánico, en definitiva, mejorará el SEO (*Search Engine Optimization*). Esto significa que un mayor número de usuarios accederán directamente a la web corporativa a través de los buscadores, ya que se estima que más del 90% de los usuarios se quedan en la primera página de resultados de Google cuando realizan búsquedas en Internet.

En el caso concreto del contexto online, la reputación se forma mediante la acumulación de opiniones o juicios que los usuarios emiten antes, durante y después del proceso de compra utilizando diferentes recursos (wikis, foros, redes, blogs, etc.). En el caso de los hoteles también son importantes las opiniones que los usuarios dejan en comparadores de precios o agencias de viajes online (OTAs). La comunicación de marketing influye en la reputación de la empresa y más en las redes sociales, que permiten una comunicación directa con los clientes. En este entorno digital es esencial mantener una rápida y efectiva respuesta y una atención personalizada a los usuarios. Además, debe tenerse en cuenta que la comunicación en el contexto de las redes sociales se focaliza en un consumidor con características destacadas: digital activo, implicado y productor de contenidos (*prosumer*).

La reputación online puede verse como la punta de lanza dentro del objetivo estratégico de las políticas de comunicación digital que aspiran a mejorar el *engagement* empresa-consumidor (Fondevila, *et al.*, 2012). Teniendo en cuenta la importancia de las redes sociales en nuestra sociedad, una buena reputación online es esencial tanto para la imagen de marca como para la reputación corporativa de cualquier empresa y, muy especialmente, en el caso de las empresas que operan en el sector turístico.

2.4.1. Herramientas para la adecuada gestión de la reputación online

Los hoteles son cada día más conscientes de la importancia de utilizar herramientas que favorezcan que sus potenciales clientes perciban una imagen positiva de los servicios que pueden ofrecer. Para lograr esa buena reputación online los hoteles pueden utilizar diferentes instrumentos:

- Herramientas de supervisión de contenidos y generación de alertas. Por ejemplo, mediante *Google Alerts* se puede automatizar la recepción de información sobre nuevo contenido en la Web, blogs, videos, que haga referencia a una serie de palabras clave. Se pueden utilizar servicios de *clipping* como *ePrensa* para rastrear la presencia de la marca en la prensa digital.
- Herramientas de monitorización de presencia web. Existen herramientas como *Brand24* que permiten estar al día de lo que se publica sobre la marca o empresa en redes sociales, blogs, páginas web, noticias, etc. Otras herramientas como *SocialBro* permiten identificar audiencias relevantes y generar estrategias en base a ello.

- Gestores de redes sociales, que permitan a la empresa planificar su estrategia de publicación de contenido en diferentes redes sociales. Un ejemplo de plataforma podría ser *Hootsuite*, que cubre múltiples plataformas como *Twitter*, *Facebook*, *LinkedIn*, *WordPress* y *Google+*. Herramientas como *TweetReach* permiten monitorear el alcance de nuestros *tweets* e identificar los seguidores con más influencia.
- Herramientas de analítica web que permitan a la empresa identificar cómo se comportan los usuarios en sus visitas a la página web, cómo interaccionan con las diferentes partes de la web corporativa, identificar perfiles de usuarios o conocer el *funnel* o embudo de conversión. Por ejemplo, *Google Analytics* permitiría conocer el perfil de usuarios que se informan sobre una reserva en la página web, pero no llegan a completarla. Servicios como *TripAdvisor* ofrecen a los hoteles herramientas analíticas.
- Herramientas de diseño para hacer infografías, imágenes personalizadas, diseños con diferentes tipografías e iconos. Un ejemplo de este tipo de herramientas puede ser *Canva*.
- Herramientas de procesamiento del lenguaje natural (PLN) que pueden ser utilizadas para automatizar la lectura de cantidades masivas de texto, identificar sentimientos a partir del análisis del texto y diferenciar de manera automática las opiniones positivas frente a las negativas.

2.4.2. El papel del *community manager*

El factor más importante para una adecuada gestión de la presencia digital y la reputación online de la empresa es el factor humano. Se requiere de profesionales con nuevos perfiles y nuevas competencias y, en este contexto, surge el concepto de *community manager*, cuyas principales funciones serían:

1. Potenciar la comunicación, logrando llegar a más personas.
2. Poseer conocimientos técnicos y tácticos para obtener mejor profundidad con la estrategia en redes sociales.
3. Hacer frente a todo tipo de dudas y problemas que planteen los clientes en el contexto digital.
4. Identificar lo que se dice de la empresa, tanto de forma positiva como negativa.
5. Identificar posibles potenciales clientes en función de las opiniones o comentarios que expresan en las redes sociales.

Como se ha dicho, el elemento clave que contribuye a formar una reputación online es el conjunto de comentarios y opiniones que los clientes dejan en los distintos medios sociales. Conocer esas opiniones es importante para las empresas, pero también para sus competidores y los propios clientes. Una adecuada gestión de las redes sociales permite a las organizaciones responder a problemas e incidentes en tiempo real antes de que se conviertan en una crisis potencial, conectar personalmente con los consumidores o clientes y usarlas como un canal más de publicidad (Ordieres *et al.*, 2017). Básicamente la labor del *community manager* es la de construir, gestionar y administrar la comunidad online alrededor de una marca en Internet, creando y manteniendo relaciones estables y duraderas con sus clientes y cualquier usuario interesado en la marca (IEBS, 2019). Hoy en día esta figura es esencial en cualquier departamento de marketing, ya que el *community manager* no solo se encarga de las redes sociales, sino que realiza otras tareas relacionadas con el marketing como las relaciones públicas, la gestión de contenidos o la publicidad.

3. METODOLOGÍA

Tomando como referencia el objetivo general de este Trabajo Fin de Grado que se orienta hacia la gestión eficaz de los comentarios y opiniones de los huéspedes de hoteles, se ha considerado oportuno proponer el uso de diferentes metodologías que permitan optimizar dicha gestión. En esta sección se describen brevemente dichas metodologías.

3.1. ANÁLISIS AUTOMATIZADO DE OPINIONES EN REDES SOCIALES

Dadas las importantes implicaciones que la reputación online puede tener en la capacidad de los hoteles en atraer a nuevos clientes o fidelizar a los ya existentes, es importante para cualquier empresa del sector ser capaz de gestionar de manera eficaz la imagen que proyectan en Internet. En particular, necesitan conocer dónde se habla de ellas y lo que se está diciendo. También es importante saber diferenciar lo que son opiniones positivas de las negativas y establecer estrategias de comunicación para contrarrestar especialmente estas últimas.

Teniendo en cuenta que la mayor parte de las opiniones suelen ser presentadas en modo textual, básicamente se requerirían dos operaciones esenciales:

1. Recopilación de los mensajes de interés segmentados según diferentes atributos (demográfico, geográfico, temporal, etc.)
2. Análisis de los sentimientos que expresa cada mensaje, para ello es preciso analizar el lenguaje del propio usuario e incluso tratar de identificar aspectos como el sarcasmo o ironía.

Para ambas operaciones existen herramientas informáticas que permiten la automatización y por tanto llegar a analizar volúmenes de opiniones que no serían posibles de analizar de manera personal.

3.2. OPINION MINING (OM)

La minería de opiniones (*Opinion Mining*) o análisis de sentimientos es una disciplina relativamente nueva que cada vez tiene mayor importancia debido a la gran cantidad de comentarios que escriben los usuarios de Internet y los cambios en los hábitos de compra de los consumidores, que cada vez se informan más antes de comprar un producto o servicio.

La extracción y procesamiento de las opiniones resulta una tarea compleja y difícil si se hace de forma manual, sobre todo por el volumen de información que se debe gestionar. Por eso resultan esenciales herramientas informáticas y computacionales que puedan automatizar estos procesos.

La OM básicamente trata de clasificar las opiniones o comentarios en función de lo que expresa su autor, utilizando técnicas de procesamiento de lenguaje natural (PLN) y de minería de texto. Busca analizar las opiniones, sentimientos, valoraciones, actitudes y emociones de las personas hacia entidades como productos, servicios, organizaciones, individuos, problemas y sucesos (Liu, 2012).

Una de las primeras aplicaciones de la OM es la determinación de la polaridad de las opiniones, clasificándolas de manera binaria en positivas o negativas, o utilizando una

escala de valoración. Por ejemplo, para realizar esta labor sobre comentarios publicados en *Twitter* existen incluso herramientas online, como podría ser la herramienta *Mr. Tuit*. La Figura 3.1 muestra ejemplos de opiniones que son clasificadas como positivas o negativas por la herramienta *Mr. Tuit*.



Figura 3.1. Ejemplos de clasificación de sentimiento de opiniones con la herramienta *Mr. Tuit*.

A la hora de abordar este problema de determinar la polaridad de las opiniones existen dos tipos de metodologías (Martínez *et al.*, 2011):

1. Aproximación basada en aprendizaje automático, que utiliza un conjunto de datos preclasificados para entrenar el algoritmo de clasificación.
2. Aproximación basada en orientación semántica, que no necesita un entrenamiento previo, sino que analiza la orientación de las palabras del comentario u opinión.

El primer enfoque puede resultar más eficaz, pero resulta más costoso y complejo debido a la necesidad de contar con un buen conjunto de datos de entrenamiento y la aplicación de técnicas de *machine learning*. El segundo enfoque puede resultar más sencillo si se dispone de los recursos lingüísticos adecuados, como diccionarios o léxicos catalogados con su sentimiento. Por ejemplo, el hecho de que en un comentario aparezcan palabras como “fenomenal”, “genial” o “disfrutar” parece señalar un sentimiento positivo por parte del usuario que publicó el comentario.

En este trabajo se ha optado por este segundo enfoque para el análisis de opiniones de huéspedes de hoteles, debido a que resulta más rápido de implementar y existen aplicaciones exitosas en otros contextos, como por ejemplo el análisis de los discursos de candidatos presidenciales en *Twitter*.

Todo sistema que trate de realizar OM debe seguir el proceso que se muestra en el Figura 3.2 y que es una adaptación del propuesto por Kumar y Reddy (2016).

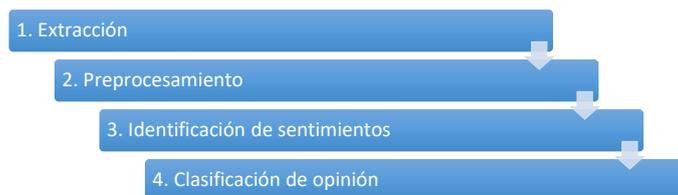


Figura 3.2. Fases de todo proceso de opinion mining.

3.3. HERRAMIENTAS UTILIZADAS

Para el desarrollo de este trabajo se han utilizado las siguientes herramientas:

- *TripAdvisor* como fuente primaria de datos, extrayéndose directamente opiniones reales de huéspedes de hoteles.
- *Brand24* para el análisis de la presencia de marca en Internet.
- *RStudio*, como entorno de desarrollo integrado para programación en *R*.
- Librerías específicas de *R* para realización de operaciones de minería de texto: *tidyverse*, *tidytext*, *tm*, *wordcloud*.
- Librerías de *R* para representaciones gráficas: *scales*, *ggplot2*, *RColorBrewer*.
- *Mr. Tuit*, herramienta para el análisis de sentimiento de *tweets*.

En las siguientes secciones se describirá con más detalle el desarrollo realizado utilizando las metodologías y herramientas descritas.

4. DESARROLLO EMPÍRICO

Con objeto de realizar un diagnóstico de cómo los hoteles abordan el tratamiento de los comentarios que sus huéspedes dejan en Internet, se ha tomado como referencia una muestra de hoteles que se dirigen a un mismo segmento de mercado. En concreto, se ha limitado este estudio a los hoteles de cuatro y cinco estrellas de la ciudad de Santander. La muestra por tanto, está formada por los diez hoteles que se muestran en la Figura 4.1.



Figura 4.1. Hoteles de la ciudad de Santander objeto de análisis.

4.1. EJEMPLO DE MONITORIZACIÓN DE MENCIONES EN INTERNET: CASO HOTEL REAL DE SANTANDER

En el caso del sector hotelero, el nivel de madurez de las herramientas para obtener información de los usuarios es mayor que en otros sectores. Las agencias de viajes online como *Booking* suelen permitir a sus clientes dejar opiniones o valoraciones de los hoteles. Por ejemplo, la Figura 4.2 muestra la valoración media de los comentarios recibidos del Hotel Real a través de esta plataforma. Como puede verse, los clientes pueden valorar diferentes servicios de hotel y los potenciales futuros huéspedes pueden incluso filtrar los comentarios basándose en diferentes parámetros. Por ejemplo, un cliente que se esté planteando alojarse en el hotel con un simple *click* puede consultar todas las opiniones sobre el desayuno del hotel, la limpieza o el trato del personal, entre otros.

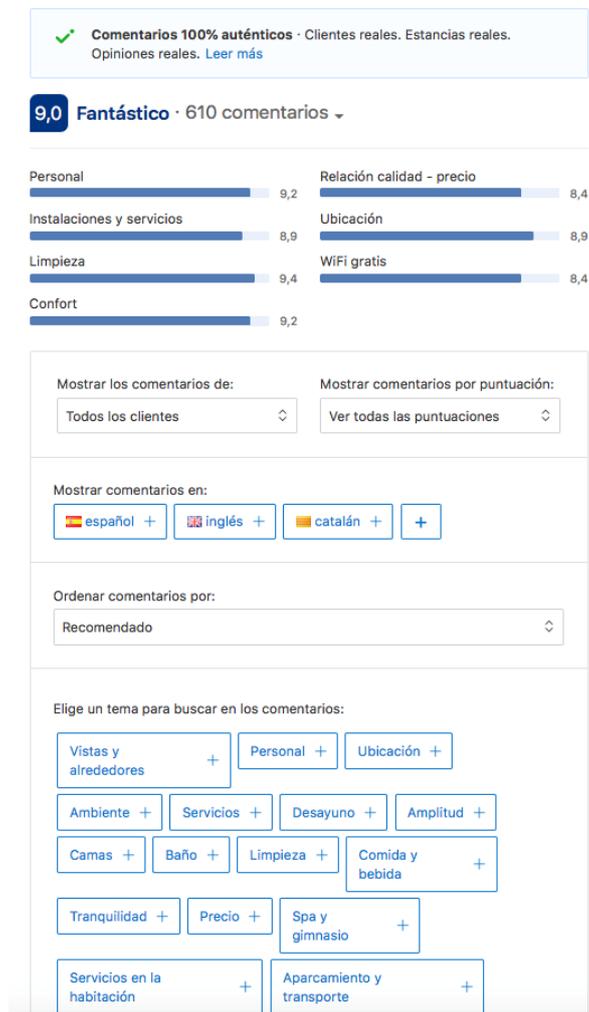


Figura 4.2. Valoraciones de clientes del Hotel Real de Santander a través de Booking.

Por otro lado, la existencia de comparadores de precios como *TripAdvisor* también permite a los clientes y a los hoteles acceder a información muy valiosa sobre las percepciones de los clientes. La Figura 4.3 muestra un ejemplo de la información que proporciona *TripAdvisor* como agregador de opiniones. En este caso, se puede decir que la información está totalmente organizada, todos los comentarios sobre un hotel se muestran juntos, con una valoración cualitativa y cuantitativa y con información sobre el perfil del usuario y del momento en el que se realiza el comentario.

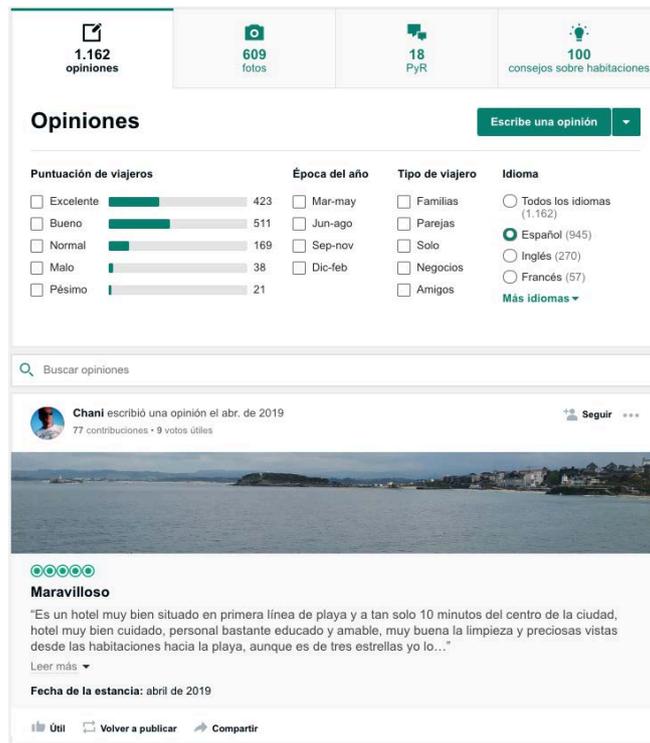


Figura 4.3. Opiniones a través de la plataforma TripAdvisor.

En cambio, en otras fuentes de información como *Facebook* o *Twitter*, también se pueden estar publicando opiniones importantes para la reputación online del hotel y se corre el riesgo de que pasen desapercibidas para la empresa. En ese caso, convendría utilizar herramientas como *Brand24* que permiten realizar al mismo tiempo búsquedas y análisis de información.

Algunas de las características básicas de *Brand24* son las siguientes:

- Permite seleccionar diferentes fuentes de información (redes sociales, blogs, noticias, etc.). Además, se puede vincular un proyecto con la página de reseñas de la marca de *Google+* y la de comentarios de *TripAdvisor*.
- Se pueden realizar filtros temporales, geográficos y según el origen de la mención.
- Realiza una clasificación de las menciones en tres categorías: positiva, neutral y negativa, de acuerdo al análisis del texto.
- Genera modelos de informes con diferentes indicadores, como por ejemplo, porcentaje de menciones positivas/negativas, número de *likes*, alcance estimado.
- Asigna un nivel de influencia (*influencer score*) a cada usuario que realiza una mención a la marca.
- Permite realizar análisis comparativos entre marcas, por ejemplo, para compararse con la competencia.

La Figura 4.4 muestra la pantalla principal de esta herramienta analizando las menciones del Hotel Real de Santander durante el periodo comprendido entre el 18/02 y el 19/05 de 2019. Precisamente, se observa una mayor presencia en la Web en el periodo inmediatamente posterior a las vacaciones de Semana Santa.

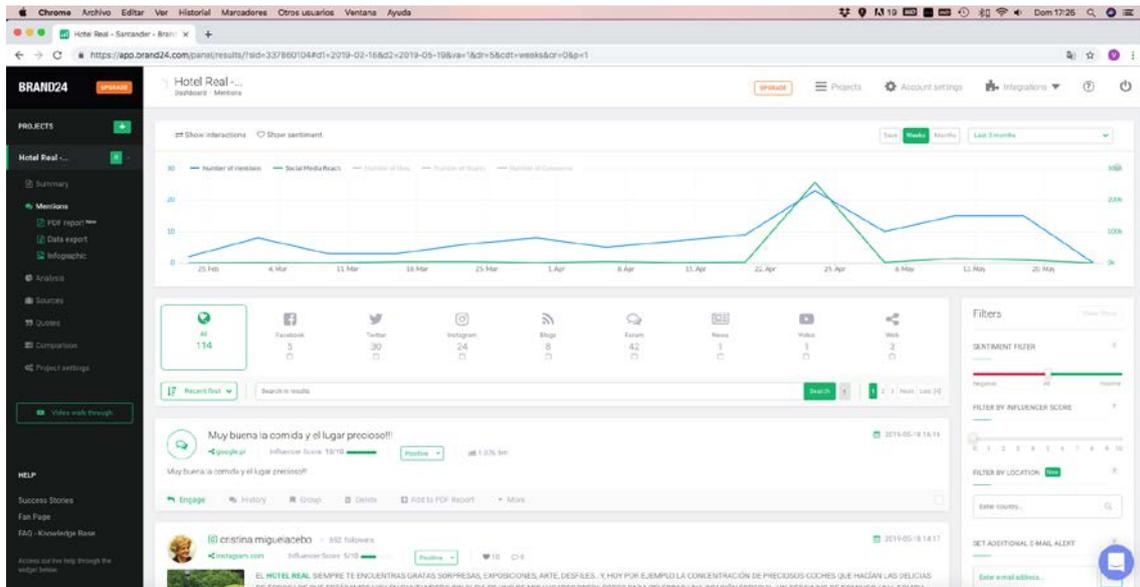


Figura 4.4. Herramienta Brand24 para la monitorización de la presencia web.

La Figura 4.5 muestra algunos indicadores importantes sobre la presencia web del Hotel Real durante ese periodo de tres meses. Por ejemplo, la herramienta estima un alcance de 300.352 usuarios que han leído alguna de las 114 menciones. También se observa un importante porcentaje de menciones positivas.



Figura 4.5. Indicadores sobre la presencia web del Hotel Real de Santander a través de la plataforma Brand24.

La Figura 4.6 muestra de manera gráfica de dónde vienen las menciones. En este caso, la mayor parte de las menciones provienen de foros (*Google+* y *TripAdvisor*), seguido de *Instagram* y *Twitter*. También se muestra gráficamente el análisis del sentimiento de las menciones. Todos estos gráficos son generados por la propia herramienta *Brand24*.

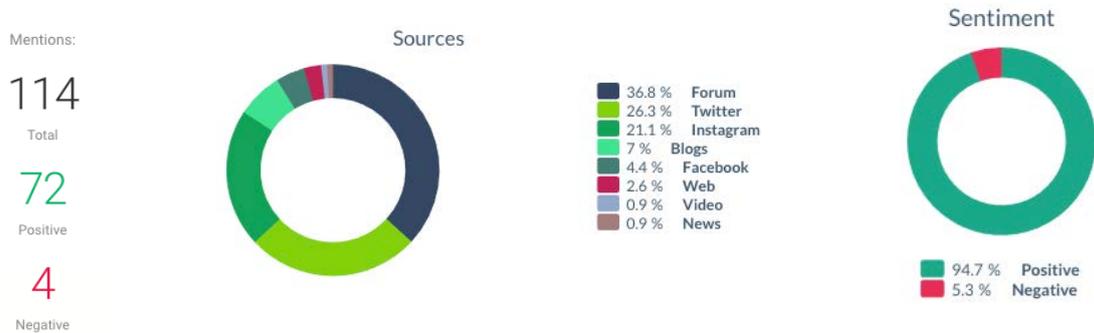


Figura 4.6. Resumen de las fuentes de información con menciones al Hotel Real de Santander.

La herramienta también permite consultar las distintas menciones y muestra para cada una de ellas de dónde proviene, una valoración de la influencia del creador del contenido, junto con la clasificación del sentimiento del comentario. Por ejemplo, la Figura 4.7 muestra un comentario positivo proveniente de *Instagram*, mientras que la Figura 4.8 muestra un comentario negativo en una reseña de *Google+*.



Figura 4.7. Ejemplo de mención calificada como "positiva" por Brand24.



Figura 4.8. Ejemplo de mención calificada como "negativa" por Brand24.

Es importante observar que este tipo de herramientas permiten monitorear la presencia de la marca no solo en redes sociales o plataformas vinculadas con el sector turístico (OTAs, comparadores de precios, etc.), sino también en otras fuentes que pueden tener mayor alcance y, sin embargo, desde el hotel pueden pasar desapercibidas. Por ejemplo, la Figura 4.9 muestra una mención positiva que aparece del hotel en una revista de la prensa rosa. Como se puede ver, el alcance de dicha publicación está próxima a los 9 millones de usuarios.



Figura 4.9. Ejemplo de mención calificada como “positiva” por Brand24 proveniente de prensa digital.

En definitiva, los hoteles deberían prestar atención a su presencia en distintas y muy variadas fuentes de información online, tratar de identificar la imagen de la marca en internet y establecer estrategias para mejorar su reputación online, contrarrestar las posibles menciones negativas y aprovechar las positivas para obtener ventajas competitivas. Para todo ello, herramientas como la que se ha mostrado resultan muy útiles.

4.2. ANÁLISIS DE SENTIMIENTO DE LOS HOTELES DE LUJO DE LA CIUDAD DE SANTANDER

Cuando un cliente deja su valoración a través de un portal como *TripAdvisor*, además de dejar su comentario suele realizar valoraciones cuantitativas sobre su experiencia global o sobre distintos aspectos del hotel. Esas valoraciones cuantitativas son más fáciles de tratar por parte de los responsables del hotel, por ejemplo, estableciendo valoraciones medias, identificando los servicios que son peor o mejor valorados, etc. En cambio, los comentarios textuales publicados en esos servicios o en redes sociales, también pueden reflejar un sentimiento, aunque a veces son más difíciles de tratar o incluso de asociarlos un valor cuantitativo que permita hacer comparaciones.

En esta sección, se realizará precisamente un análisis de sentimiento de los comentarios textuales que reciben los hoteles objeto de análisis en este trabajo, es decir, los hoteles de cuatro y cinco estrellas de la ciudad de Santander. Para ello, se tomarán como referencia las etapas básicas de todo proceso de OM que se mostraron en la Figura 3.2. En las próximas subsecciones se describen las tareas básicas realizadas.

4.2.1. Extracción de opiniones

La primera fase consiste en extraer el texto que contiene las opiniones. Esto puede hacerse recuperando opiniones de la Web, redes sociales, blogs, etc. Para esta labor existen diferentes herramientas, por ejemplo, *Twitter* ofrece a los usuarios una API (*Application Programming Interface*) que permite conectar *Twitter* con otros programas como *R* o *Python*, dos de los lenguajes más utilizados en procesos de Big Data. Gracias a esta API y a paquetes como *twitteR* se podría programar de manera automática la recuperación de todos los *tweets* a partir de una fecha determinada y que contengan determinado *hashtag* o un texto en su contenido.

Otra de las opciones es la utilización de técnicas de *web scrapping*. Estas técnicas permiten extraer información de páginas web de manera automatizada, analizando el código fuente que hay detrás de cada página. Un ejemplo podría ser el paquete *Rvest* del software *R*.

Para obtener un conjunto de opiniones que puedan ser utilizadas para realizar OM sobre los hoteles de Santander, se optó por extraer directamente comentarios de la página de *TripAdvisor*. Para acotar el número de opiniones se extrajeron únicamente

las correspondientes a un periodo de un año a partir de junio de 2018. Se recuperaron un total de 698 opiniones sobre el total de los 10 hoteles que forman la muestra. Para cada una de las opiniones fue recuperada la fecha de publicación, el contenido del comentario y la valoración del usuario en la escala 1-5 que *TripAdvisor* ofrece.

La Tabla 4.1 indica el número de opiniones que se extrajeron para cada hotel, en el periodo junio 2018 – mayo 2019, junto con las valoraciones medias dadas por los huéspedes en el portal *TripAdvisor*.

Hotel	Número de opiniones	Valoración media (1-5)	Desviación típica
BAHIA	114	4.45	0.810
COLISEUM	49	4.12	0.881
HOYUELA	19	4.00	1.20
PALACIO-MAR	26	3.35	1.32
REAL	68	4.10	1.22
RIO	80	3.65	1.22
SANTEMAR	178	4.09	0.878
SARDINERO	53	4.32	0.754
VICTORIA	43	3.53	1.64
VINCCI	68	4.25	1.04

Tabla 4.1. Resumen de las opiniones extraídas sobre hoteles 4 y 5* de Santander.

En términos globales, se observa que la valoración media de los hoteles es muy buena, ya que el 70% de ellos tiene una valoración superior a 4. Este primer análisis parece mostrar que en general los hoteles de 4 y 5 estrellas de la ciudad de Santander vienen a cumplir las expectativas de los clientes que optan por este tipo de alojamientos.

4.2.2. Preprocesamiento

Con todas las opiniones se generó un archivo CSV exportado desde *Excel* en el que cada fila representa un comentario e incluye los siguientes elementos separados por comas:

- Un identificador numérico de cada comentario
- Nombre del hotel al que hace referencia
- Fecha de publicación (mes y año)
- Texto con el contenido del comentario
- Valoración del huésped (escala 1 – 5)

Para facilitar el procesamiento posterior se realizaron las siguientes operaciones:

- Todo el texto fue convertido a letras minúsculas
- Se eliminaron espacios en blanco duplicados, cambios de línea, signos de puntuación y algunos caracteres especiales (comillas, interrogaciones, guiones, etc.).
- Se eliminaron algunas palabras que resultan muy frecuentes (como la palabra “no”) y no aportan ninguna información sobre el mensaje, las que se conocen como *stopwords*.

A modo de ejemplo, la Figura 4.10 muestra cómo se refleja un comentario real en el fichero final de datos tras realizar este proceso.

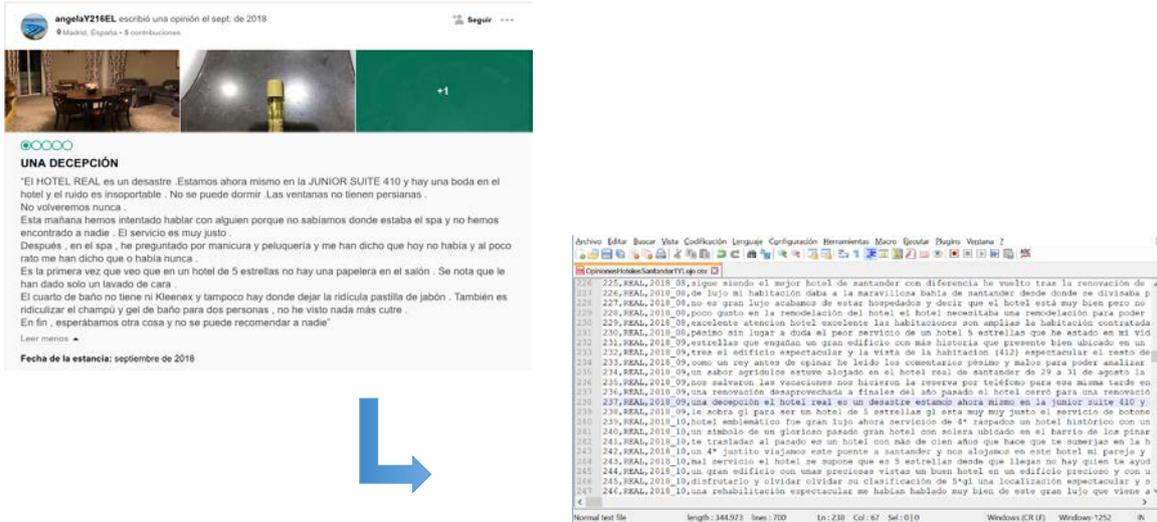


Figura 4.10. Preprocesamiento de las opiniones.

La fase de preprocesamiento finalizó separando individualmente cada palabra de cada comentario, pero guardando el identificador del comentario para poder saber en todo momento a qué opinión pertenece cada palabra. Esto se realizó con el software *R*.

4.2.3. Identificación de sentimientos

Para la identificación de sentimientos se ha utilizado también el software *R* y en concreto los paquetes *tidytext*, *tidyverse* y *tm*, que contienen herramientas para la manipulación de texto. El paquete *tidytext* incorpora un léxico de palabras, *Afinn*, con puntuaciones que oscilan entre -5 y 5 según su percepción negativa o positiva. La versión original del léxico *Afinn* está en inglés, pero en este caso se usó la traducción aportada por Juan Bosco Mendoza¹.

El léxico incluye 2.476 palabras y aunque es de carácter general y tiene algunas valoraciones que pueden ser discutibles, es suficiente para lograr los objetivos de este trabajo. La Tabla 4.2 muestra algunos ejemplos de palabras incluidas en ese diccionario.

Palabras con percepción negativa		Palabras con percepción positiva	
Puntuación	Ejemplos	Puntuación	Ejemplos
-5	puto, coño	5	encantado, excepcional
-4	mierda, fraude	4	fabuloso, triunfo
-3	furioso, harto	3	vibrante, admirar
-2	ignorado, inaceptable	2	afortunado, agradable
-1	manipulación, olvidado	1	atraer, legal

Tabla 4.2. Ejemplos de palabras del léxico *Afinn*.

Para el análisis de sentimientos, se optó por una aproximación basada en orientación semántica, que no necesita un entrenamiento previo, sino que analiza directamente la orientación de las palabras del comentario u opinión. Con este enfoque, el sentimiento de un comentario podría calcularse agregando las puntuaciones de las palabras que

¹ https://rpubs.com/jboscomendoza/analisis_sentimientos_lexico_afinn

En definitiva, de cada comentario se podrían obtener dos valores:

1. Valoración directa del huésped en la escala 1 – 5
2. Puntuación OM en la escala -5 – 5

Ambos valores, de alguna manera servirían para identificar el sentimiento del comentario.

4.2.4. Clasificación de opinión

Sobre las opiniones recuperadas de los hoteles se calculó la media aritmética de puntuaciones de las palabras. De esta manera, cada opinión recibió una puntuación global, siendo la media de las 698 opiniones la puntuación de 1.19. Esto parece indicar que en general la valoración de los huéspedes de los hoteles de lujo de la ciudad de Santander es positiva.

Si se analizan las puntuaciones medias de las opiniones de cada hotel y se comparan con las valoraciones directas del propio huésped en *TripAdvisor*, se obtienen los datos que se muestran en la Tabla 4.3.

Hotel	Número opiniones	Valoraciones directas del huésped		Puntuaciones automáticas OM	
		Media	Desviación típica	Media	Desviación típica
BAHIA	114	4.45	0.810	1.43	1.40
COLISEUM	49	4.12	0.881	1.12	1.72
HOYUELA	19	4.00	1.20	0.996	1.54
PALACIO-MAR	26	3.35	1.32	0.951	1.68
REAL	68	4.10	1.22	0.886	1.59
RIO	80	3.65	1.22	1.02	1.48
SANTEMAR	178	4.09	0.878	1.20	1.62
SARDINERO	53	4.32	0.754	1.59	1.24
VICTORIA	43	3.53	1.64	0.909	1.78
VINCCI	68	4.25	1.04	1.31	1.63

Tabla 4.3. Comparación entre las valoraciones directas de los huéspedes y las puntuaciones obtenidas con técnicas Opinion Mining (OM).

En la tabla anterior, las celdas sombreadas en verde representan los hoteles con mejor valoración o puntuación; mientras que las celdas en rojo muestran los hoteles con peores resultados. Analizando esos datos, se podría decir que los clientes suelen ser coherentes a la hora de dar su valoración numérica y que comentarios que incluyen palabras de percepción negativa se corresponden a malas valoraciones en *TripAdvisor*.

De manera gráfica, se observa una clara correlación entre las puntuaciones OM y las valoraciones *TripAdvisor*, tal como muestra la Figura 4.12.

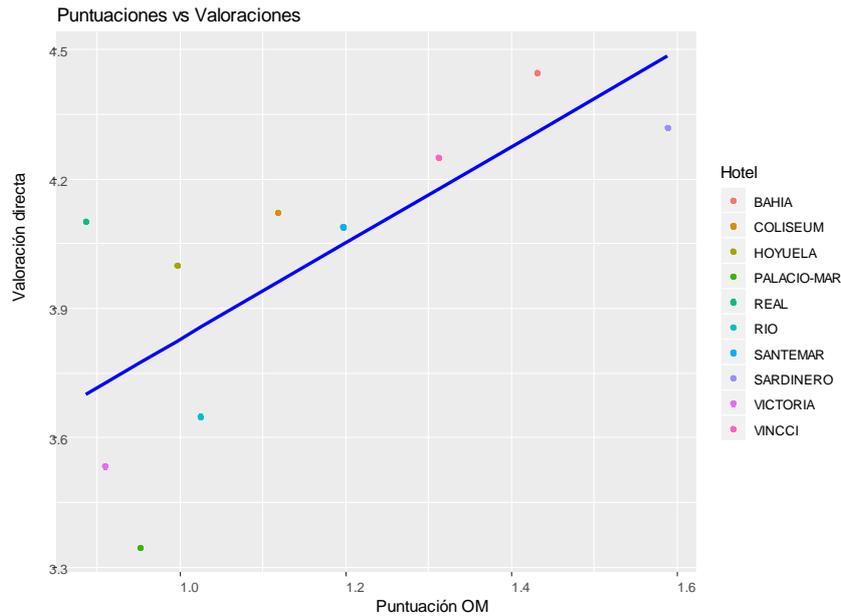


Figura 4.12. Correlación positiva entre puntuaciones OM y valoraciones directas en *TripAdvisor*.

Haciendo uso del software *R*, puede obtenerse también la ecuación de regresión que estimaría la valoración del hotel en *TripAdvisor* a partir de la puntuación obtenida al aplicar OM sobre los comentarios de los usuarios. La ecuación de regresión obtenida sería la siguiente:

$$\text{valoración} = 2.7122 + 1.1167 * \text{puntuación}$$

La recta de regresión obtenida parece indicarnos que como punto de partida la valoración de los hoteles es positiva (2.7122), y que por cada punto positivo que obtenga el análisis de sentimientos del comentario del huésped, se incrementará en más de un punto la valoración directa estimada en *TripAdvisor*. Por ejemplo, en el comentario de la Figura 4.11, la valoración que se estimaría es:

$$\text{valoración} = 2.7122 + 1.1167 * 1,545 = 4,44$$

Este valor es coherente con la valoración de 5 que el huésped dio directamente en *TripAdvisor*.

Este primer análisis parece confirmar que un análisis directo de los mensajes de los clientes es un buen indicador de la valoración que el propio huésped pueda dar en sistemas como *TripAdvisor*. Por tanto, esta metodología también se podría aplicar sobre comentarios que los clientes publican en otras plataformas en las que no se realizan valoraciones cuantitativas, por ejemplo, en redes sociales como *Twitter*, *Facebook*, *Instagram*, incluso en blogs o foros.

Para toda empresa es importante conocer y valorar su reputación online pero también la de sus competidores. Un análisis comparativo de este tipo puede servir para identificar posibles ventajas competitivas. Para un hotel también es importante realizar este tipo de análisis. Las técnicas de *opinion mining* pueden facilitar la comparación del sentimiento que los clientes trasladan en el entorno digital. Por ejemplo, la Figura 4.13 muestra mediante un *boxplot* las distribuciones de las puntuaciones obtenidas por los diez hoteles analizados en este estudio. Como puede verse, a pesar de que el Hotel Real es a priori el más lujoso y de más prestigio, la puntuación media de las opiniones es la más baja. Quizás esto se deba a que los clientes de un hotel de cinco estrellas tienen mayores expectativas y mayor grado de exigencia.

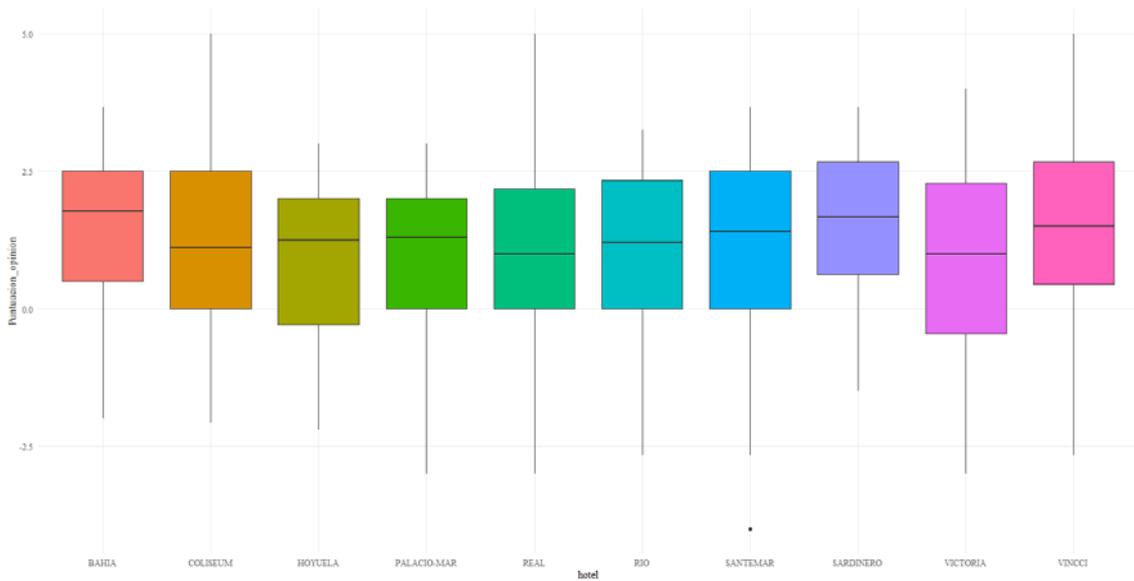


Figura 4.13. Puntuaciones medias obtenidas por los hoteles de lujo de Santander.

Como se ha comentado anteriormente, se consideran como opiniones con sentimiento positivo aquellas que obtienen puntuaciones mayores que 0, mientras que se consideran negativas las que tienen puntuaciones por debajo de 0. En la Figura 4.14 pueden verse las diferentes proporciones entre opiniones positivas y negativas en cada uno de los hoteles. Como puede comprobarse, dos hoteles tienen una mejor proporción de comentarios positivos: el Hotel Bahía y el Hotel Sardinero. En todos los hoteles los comentarios positivos dominan sobre los negativos.

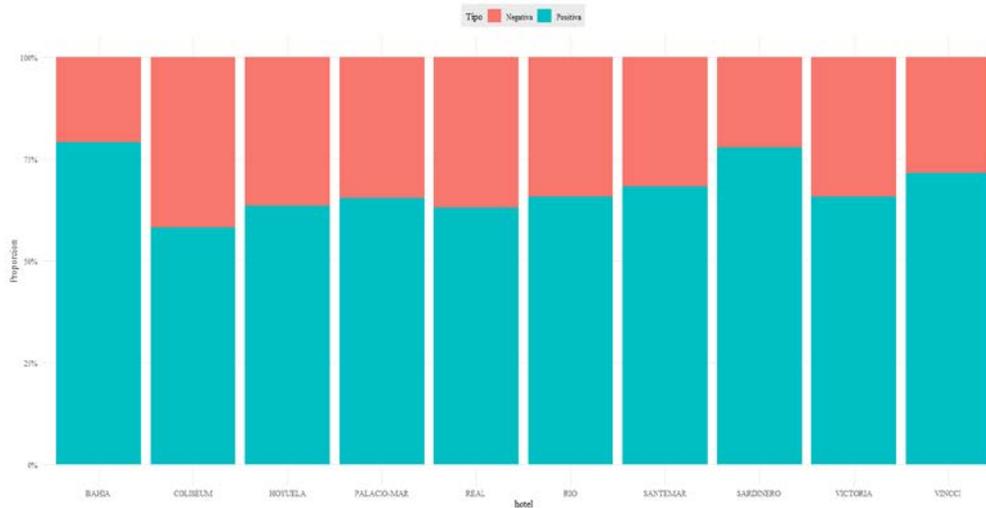


Figura 4.14. Proporción de opiniones con sentimiento positivo/negativo en los hoteles analizados.

Finalmente, otro estudio interesante es identificar las palabras con connotaciones positivas y negativas más frecuentes en las opiniones que recibe cada hotel. Especialmente las palabras negativas pueden facilitar la identificación de problemas en la gestión del hotel y poder ayudar a tomar las decisiones adecuadas. Las Figuras 4.15 y 4.16 muestran las palabras positivas y negativas más frecuentes en las opiniones de los diez hoteles y el número de opiniones en las que la respectiva palabra aparece.

Aplicación de técnicas de Opinion Mining para el análisis de la reputación online en el sector hotelero



Figura 4.15. Palabras positivas más frecuentes en las opiniones de los hoteles.



Figura 4.16. Palabras negativas más frecuentes en las opiniones de los hoteles.



Figura 4.18. Palabras más frecuentes en los comentarios negativos sobre el Hotel Real.

Como se puede observar en las nubes de palabras, algunas de las palabras más frecuentes en los comentarios positivos del Hotel Real son “hotel”, “desayuno”, “vistas”, “personal”, “habitaciones”, etc. Esto es un indicador claro de que los huéspedes de este hotel valoran positivamente y dan mucha importancia a factores como las vistas desde el balcón del hotel, el desayuno buffet que se oferta, o el trato tan agradable del personal. Por otro lado, fijándose en las palabras más frecuentes en los comentarios negativos, se observa una mayor existencia de menciones a palabras como “hotel”, “estrellas”, “habitación”, etc. Esto podría indicarnos que quizás los clientes ponen en duda las cinco estrellas del hotel, o que las habitaciones no sean lo suficientemente buenas para los estándares de calidad que el cliente espera.

En definitiva, este tipo de información es realmente útil para cualquier hotel. Mediante este tipo de análisis se pueden sacar conclusiones rápidas sobre aquellos factores que el cliente da mayor importancia, y aquellos que generan mayor insatisfacción y en consecuencia generan opiniones negativas en Internet. De esta manera, un hotel como el Hotel Real podrá establecer estrategias empresariales orientadas a mejorar y contrarrestar aquellos aspectos peor valorados por los clientes y fortalecer aquellos más positivos.

4.3. RESPUESTAS DE LOS HOTELES ANTE LOS COMENTARIOS DE LOS HUÉSPEDES

Para terminar el análisis empírico es interesante analizar cómo los hoteles responden o reaccionan ante los comentarios de los huéspedes. Aunque aún hay pocos trabajos que realicen estudios sobre las respuestas de los hoteles a los comentarios de los clientes, algún trabajo pone de manifiesto que son minoría los hoteles que entran a responder a esos comentarios publicados en *TripAdvisor*. Por ejemplo, un reciente estudio sobre hoteles de Madrid y París pone de manifiesto que el 41,5% de los hoteles responde a los comentarios online generados entre el 80% y el 100% de las ocasiones, aunque solo un 14,9% lo hace siempre (San Martín *et al.*, 2018). Ese estudio también pone de manifiesto que los hoteles suelen responder con mensajes

predefinidos y que únicamente personalizan los comentarios en las respuestas al 40,27% de los comentarios positivos y el 37,43% de los negativos.

Con objeto de valorar si esos resultados son también aplicables a la muestra de hoteles analizados en este trabajo, se comprobó cuantas respuestas habían publicado los hoteles ante los 698 comentarios de los usuarios. En concreto, las tasas de respuesta fueron las que se muestran en la Tabla 4.4. Como puede verse, solo 3 hoteles (Coliseum, Rio y Santemar) responden prácticamente a todos los clientes, agradeciendo sus valoraciones positivas o disculpándose por los inconvenientes que los usuarios trasladan. Normalmente en esas respuestas recurren a fórmulas genéricas. En uno de los hoteles (Victoria) se trató de dar respuesta a un comentario que resultaba especialmente negativo. Las tasas de respuesta son ligeramente más bajas que las correspondientes a los hoteles de Madrid y Paris analizados en el estudio anterior.

Hotel	Opiniones	Respuestas	%
BAHIA	114	0	0,00%
COLISEUM	49	47	95,92%
HOYUELA	19	0	0,00%
PALACIO-MAR	26	0	0,00%
REAL	68	0	0,00%
RIO	80	78	97,50%
SANTEMAR	178	175	98,31%
SARDINERO	53	0	0,00%
VICTORIA	43	1	2,33%
VINCCI	68	0	0,00%

Tabla 4.4. Tasas de respuesta de los hoteles de Santander a los comentarios de sus usuarios.

Para gestionar adecuadamente las respuestas existen herramientas informáticas de ayuda. Estas herramientas podrían ser utilizadas por los hoteles para atender mejor a los clientes en redes sociales y entornos de interacción social. Por ejemplo, una herramienta como *HootSuite* permite administrar el contenido, entre otras, de las siguientes redes sociales: *Facebook*, *Twitter*, *LinkedIn*, *Google*, *Instagram*, *YouTube*. Esta herramienta facilita la búsqueda, programación, gestión y análisis de la efectividad del contenido de la empresa en redes sociales.

5. CONCLUSIONES

En la sociedad actual, caracterizada por el uso cotidiano de las tecnologías de la información y las comunicaciones y la facilidad para la interacción social en contextos globales, resulta esencial para las empresas tener una imagen fiel de su reputación y de cómo sus clientes perciben su imagen de marca. En el sector turístico, y en particular en el hotelero, esto es especialmente importante debido al mayor nivel de madurez y de desarrollo del comercio electrónico y a los hábitos de compra de los consumidores. En este contexto, el consumidor se convierte en un productor de contenidos (*prosumer*) y al mismo tiempo cada vez más utiliza diversas fuentes de información antes de realizar una reserva de alojamiento o adquirir otro producto turístico. Entre las fuentes de información, las redes sociales, los agregadores de opiniones, los comparadores de precios e incluso blogs y foros de debate, son cada vez más utilizados por una parte importante de los consumidores. Toda esta información constituye la reputación online de la empresa y la misma debería preocuparse por gestionarla de manera adecuada y definir estrategias que mejorarla.

Un elemento esencial para esta gestión adecuada es el uso de una combinación de herramientas y metodologías que permitan conocer en todo momento lo que se dice de la empresa, a través de qué medios y sobretodo ser capaz de identificar posibles clientes descontentos, con sentimiento negativo hacia la marca y poder analizar las causas de ese descontento. Las redes sociales dan a los clientes un enorme poder, por lo que es esencial poder identificar las incidencias prácticamente en tiempo real y poder tomar medidas para poder contrarrestar los efectos de las críticas negativas, pero también aprovechar los efectos de las positivas. El concepto de *influencer* y su cada vez mayor relevancia es un ejemplo que pone de manifiesto el poder que determinados consumidores pueden tener a través de las redes sociales. De hecho, las empresas cada vez más utilizan estrategias de marketing basadas en conseguir que personas con un alto nivel de seguidores puedan realizar valoraciones positivas de sus productos (*marketing de influencers*).

En la actualidad existen importantes herramientas que facilitan la gestión eficiente de la reputación online, especialmente cuando se tenga que gestionar cantidades enormes de información proveniente de diversas fuentes. Por ello, las empresas cada vez más incorporan a su plantilla personas con un perfil mixto con conocimientos de comunicación, marketing y tecnología para gestionar la presencia web de la empresa y las relaciones con sus clientes a través de las redes sociales. El papel del *community manager* es cada vez más frecuente en las empresas, y en particular en las que operan en el sector turístico. Un buen *community manager* deber ser capaz de dominar herramientas de seguimiento y monitorización de la presencia web, de *business analytics*, de análisis de sentimiento, etc.

En concreto, con este trabajo se ha pretendido mostrar cómo el uso técnicas de procesamiento de lenguaje natural facilita la identificación del sentimiento de los clientes de hoteles. En determinadas fuentes, como agregadores de opiniones o centrales de reservas online el cliente muchas veces deja una valoración cuantitativa de su experiencia en el hotel. Sin embargo, en otras ocasiones, como por ejemplo publicaciones en redes sociales, el cliente se limita a publicar un comentario describiendo su experiencia. En este último caso las técnicas de *opinion mining* pueden ser especialmente útiles para que el hotel obtenga de manera automática información sobre percepciones positivas y negativas de sus clientes. Para mostrar todo esto, en el trabajo se ha realizado un análisis de los comentarios reales en *TripAdvisor* sobre los hoteles de cuatro y cinco estrellas de la ciudad de Santander en el último año. Este trabajo ha servido para lograr los objetivos inicialmente planteados, y especialmente para mostrar la utilidad de este tipo de metodologías.

6. LIMITACIONES Y LÍNEAS FUTURAS DE TRABAJO

Este estudio tiene algunas limitaciones que deberían tenerse en cuenta para realizar futuras mejoras. En primer lugar, podría ampliarse la base de datos de comentarios, incluyendo un mayor número de comentarios, ampliando el segmento de hoteles e incorporando comentarios de distintas fuentes. Incluso podría ser interesante poder analizar de manera automática otro tipo de información no textual, por ejemplo, videos o fotos que puedan publicar los huéspedes. Obviamente esto último aumentaría la complejidad del trabajo a realizar.

Por otro lado, teniendo en cuenta la globalización del sector hotelero, sería importante poder analizar simultáneamente comentarios escritos en diferentes idiomas. En relación a esto, existen técnicas de minería de textos multilingüe que podrían ser utilizadas.

Para el análisis de sentimiento de opiniones se ha utilizado un léxico con palabras de percepción positiva y negativa. Ese léxico es de carácter general y sería interesante ampliarle con palabras, términos o expresiones específicas del sector turístico.

Una de las dificultades del análisis textual de manera automática es detectar la ironía o el sarcasmo en los comentarios. En ocasiones un comentario puede contener palabras de connotación positiva pero que en realidad reflejan un descontento. Detectar este doble sentido de forma automática no es fácil.

Finalmente, cuando se pretenden analizar opiniones de clientes en redes sociales sobre productos o servicios, surgen una serie de preocupaciones. Entre ellas se podrían citar la protección de la identidad, la posibilidad de engaño, invención de perfiles, estrategias de boicot de la competencia, *free riders* destructivos, etc. Todo ello también debe ser tenido en cuenta al abordar este tipo de tareas.

REFERENCIAS

Figuerola, M. (2018). APROXIMACIÓN A UN SISTEMA DE CUENTAS NACIONALES DEL TURISMO ESPAÑOL. Mesa del Turismo. <https://mesadelturismo.org/wp-content/uploads/2018/10/AproximacionaunsistemadecuentasnacionalesProyeccion2018.pdf>

Filleri, R.; Algezai, S.; McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51, 174-185.

Fombrun, C. J.; Gardberg, N. A.; Sever, J. M. (2000). The Reputation QuotientSM: A multi-stakeholder measure of corporate reputation. *Journal of brand management*, 7(4), 241-255.

Fondevila, J.F.; Del Olmo, J.L.; Bravo, V. (2012). Presencia y reputación digital en social media: comparativa en el sector de la moda. *Fonseca, Journal of Communication*, n.5, pp.92-116.

Fundación Cetelem (2018). El Observatorio Cetelem 2018 e-commerce. Sector viajes. <http://www.elobservatoriocetelem.es/>

Fundación Orange (2016). La transformación digital en el sector turístico. www.fundacionorange.es

Fundación Telefónica (2019). Sociedad Digital en España 2018. ISBN: 978-84-306-2311-2. www.fundaciontelefonica.com/publicaciones.

Herrero, A.; San Martín, H.; Hernández, J.M. (2015). How online search behavior is influenced by user-generated content on review websites and hotel interactive websites. *International Journal of Contemporary Hospitality Management*, 27(7), pp.1573-1597.

IEBS (2019). Guía del Community Manager. Aprende a gestionar la imagen corporativa en Internet. IEBS Business School. <https://www.iebschool.com/recursos/guia-community-manager/>

INE (2018). Cuenta Satélite del Turismo en España 2017. Publicado Dic 2018. http://www.ine.es/dynqs/INEbase/es/operacion.htm?c=estadistica_C&cid=1254736169169&menu=ultiDatos&idp=1254735576863

INE (2018b). España en Cifras 2018. Catálogo de publicaciones oficiales de la Administración General del Estado. http://www.ine.es/prodyser/espaa_cifras (ISSN 2255-0410)

Kane, G.C.; Palmer, D.; Phillips, A.N.; Kiron, D.; Buckley, N. (2015). Strategy, not technology, drives digital transformation. *MIT Sloan Management Review*. Deloitte University Press.

Kumar, B.S.; Reddy, B.S. (2016). An Analysis on Opinion Mining: Techniques and Tools. *Paripex – Indian Journal of Research*, 5(8), pp 489-492.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Martinez, E.; Martín, M.T.; Perea, J.M.; Ureña, L.A. (2011). Técnicas de clasificación de opiniones aplicadas a un corpus en español. Revista de Procesamiento del Lenguaje Natural, nº 47, pp. 163-170.

ONTSI (2018). Informe e-Pyme 2017 de análisis sectorial de la implantación de las TIC en las empresas españolas. Observatorio Nacional de las Telecomunicación y la SI (ONTSI). Red.es. Ministerio de Economía y Empresa. DOI: doi.org/10.30923/2341-4030-2018

Ordieres-Meré, J.; Franco Riquelme, J.N. (2017). Análisis de Opinión en Redes Sociales y Valor para el análisis de producto. Revista Economía Industrial, n. 405; pp. 65-74.

Pabel, A.; Prideaux, B. (2016). Social Media use in pre-trip planning by tourists visiting a small regional leisure destination. Journal of Vacation Marketing, 22(4), pp.335-348.

Rubio, A.; Jiménez, I.C.; Mercado, C. (2017). Reputación corporativa online en la hotelería: el caso de TripAdvisor. ESIC Market Economics and Business Journal, 48(3), 595-608.

San Martín, S; Jiménez, N.; Puente, N. (2018). Análisis con big data de las respuestas de los hoteles en Tripadvisor. ESIC Market Economics and Business Journal, 49(2), 359-378.

TripAdvisor (2018). TripBarometer 2017/18. Global Report. <https://www.tripadvisor.com/TripAdvisorInsights/tripbarometer>

Truxel, E. (2016). The American Traveler's Path to Purchase. [PowerPoint Deck]. Expedia Media Solutions and ComScore. Retrieved from <http://expediaslides.com/view/new/mail?ilD=SqVw3dtK7MtMcGfXvM9R>

Walsh, G.; Mitchell, V.W.; Jackson, P.; Beatty, S. (2009). Examining the antecedents and consequences of corporate reputation: a customer perspective. British Journal of Management, 20(2), 187-203.