



FACULTAD DE DERECHO

SOCIAL LISTENING: ESTUDIO DE LA REPUTACIÓN ONLINE DE LUIS RUBIALES

Autor: Beatriz Ibáñez Puerto

Director: Antonio Tena Blázquez

Madrid | Junio, 2024

Resumen

En el presente trabajo se comenzará explicando de manera teórica la terminología en torno a la ciencia de la información y la participación de las nuevas tecnologías en el tratamiento de datos, para subsiguientemente dar paso a su aplicación práctica. Para ello, se realizará un pormenorizado análisis de las conclusiones obtenidas al utilizar técnicas de *data mining*, como el *text mining* y el posterior *sentiment analysis* en RStudio, sobre un conjunto desestructurado de datos extraído de Twitter que constará de todas las opiniones de los usuarios con respecto a la polémica acción de Luis Rubiales en la entrega de premios de la final del Mundial de Fútbol Femenino.

Abstract

This paper will begin with a theoretical explanation of the terminology surrounding data science and the participation of new technologies in the treatment of raw data, in order to make its way to the practical application. To do so, data mining techniques, such as text mining and sentiment analysis, will be applied to an unstructured set of data extracted from Twitter composed of all of the opinions from media users regarding the controversial action of Luis Rubiales in the awards ceremony at the final of the FIFA Women's World Cup. Then, a thorough analysis will be carried out of the conclusions obtained from this study.

Palabras clave: Luis Rubiales, escucha social, análisis de sentimiento, Twitter, emociones.

Keywords: Luis Rubiales, social listening, sentiment analysis, Twitter, emotions.

ÍNDICE

I. Justificación del tema	6
II. Objetivos del trabajo	7
III. Marketing y publicidad en la era digital	7
A. ¿Qué es el marketing digital?	7
B. ATL versus BTL	8
C. BTL, el favorito de las empresas	10
IV. Introducción a la inteligencia de datos	10
A. Concepto	10
B. Data intelligence en relación con data mining y machine learning	11
C. Evolución histórica de la inteligencia de los datos	13
D. La inteligencia de datos para la empresa: business intelligence	14
E. La opinión pública a examen: la inteligencia de medios	15
F. La Web 2.0	17
V. Herramienta de escucha social: el social listening	17
A. Definición	17
B. El social listening integra el social media management	17
C. Social media monitoring versus social listening	18
D. Aplicaciones a la ciencia del marketing digital	19
E. Procedimiento	22
i) Identificación de objetivos y métricas	23
ii) Selección de palabras clave y plataformas	24
iii) Recopilación de datos	25
iv) Preprocesamiento de datos	26
v) Procesamiento de datos	29
vi) Análisis de datos: sentiment analysis	30
VI. El “caso Rubiales”	31
A. Contexto	31
B. Utilidad para la interpretación de la opinión pública	32
C. Metodología	33
VII. Conclusiones	34
A. Hallazgos de la investigación	34
B. Futuras líneas de investigación	42

VIII.	Bibliografia	43
IX.	Anexo	48

ÍNDICE DE FIGURAS

Figura 1: Inversión publicitaria en los medios tradicionales	9
Figura 2: Proceso KDD	13
Figura 3: Conceptos y tecnologías del <i>business intelligence</i>	15
Figura 4: Tweet Ryanair España	21
Figura 5: Fórmula Term Frequency – Inverse Document Frequency	28
Figura 6: Nube de palabras periodo 1	35
Figura 7: Nube de palabras periodo 2	35
Figura 8: Nube de palabras periodo 3	36
Figura 9: Clasificación por sentimientos del léxico NRC periodo 1	37
Figura 10: Barplot de clasificación por sentimientos del léxico NRC periodo 1	37
Figura 11: Clasificación por sentimientos del léxico NRC periodo 2	37
Figura 12: Barplot de clasificación por sentimientos del léxico NRC periodo 2	38
Figura 13: Clasificación por sentimientos del léxico NRC periodo 3	38
Figura 14: Barplot de clasificación por sentimientos del léxico NRC periodo 3	39
Figura 15: Frecuencia de las palabras en base a su polaridad con AFINN P1	40
Figura 16: Barplot del porcentaje de palabras clasificadas según AFINN en el P1	40
Figura 17: Frecuencia de las palabras en base a su polaridad con AFINN P2	41
Figura 18: Barplot del porcentaje de palabras clasificadas según AFINN en el P2	41
Figura 19: Frecuencia de las palabras en base a su polaridad con AFINN P3	41
Figura 20: Barplot del porcentaje de palabras clasificadas según AFINN en el P2	42

I. Justificación del tema

3 horas, 4 horas, 5 horas, pudiendo rozar temerosamente las 6 horas diarias de exposición a pantallas de dispositivos móviles, TikTok, en Twitter (o mejor dicho X), en Instagram, en Facebook (o sea, Meta)...según App Annie, una media de 4,8 horas al día, estudio en el que se incluye a todas las edades. Todo esto tiene que servir para algo. Nuestra adicción al constante bombardeo de información triste, inquietante, motivadora, alegre, impactante, información alarmante, emotiva, confusa, es la mejor arma de venta de las empresas. Y a la vez que consumimos, creamos, cuando comentamos en una publicación o cuando escribimos un tweet. No fue hasta hace unos pocos años, sin embargo, que comenzó a valorarse la riqueza de la información cruda y aleatoria que existe en las redes sociales, que se va acumulando y que va a permanecer siempre.

Van continuamente apareciendo nuevas herramientas de almacenamiento, gestión, organización, tratamiento y análisis de esos datos que permiten sacarles el máximo provecho. En este trabajo se hará un recorrido por algunos de estos métodos, abordándose todos los conceptos que los rodean. Desde 2012, con Google a la cabeza del control de inteligencias artificiales (tras contratar a Geoffrey Hinton¹), el crecimiento y aprovechamiento de estas ha sido vertiginoso, y caso inusual sería que, una compañía de envergadura considerable, no se sirviese de aquellas para impulsar sus estrategias de negocio.

Con el fin de estudiar el conocido como “caso Rubiales”, se recurrirá al *text mining* y posterior *sentiment analysis* de las opiniones de la sociedad virtual. Todo el mundo está a favor, todo el mundo está en contra, todo el mundo tiene una opinión que baila cada vez que sale un titular nuevo sobre el tema. El icónico “piquito” de Luis Rubiales, expresidente de la Real Federación Española de Fútbol, no solo ha retumbado en redes y afectado personalmente al susodicho – causando su despido e imputación judicial-, sino que ha sido utilizado como bandera feminista por parte de figuras públicas y empresas que buscan el aplauso de la opinión popular y, de paso, sacar beneficio de aquello. Así pues, a través del *social listening*, rastreamos las redes sociales en busca de todas las opiniones de los usuarios durante el periodo de máximo apogeo del tema, clasificándolas según su tinte más o menos positivo, negativo o neutro, y desembocaremos en hallazgos

¹ Es informático británico cuya empresa fue adquirida en 2012 por el gigante tecnológico Google tras inventar la primera IA.

patrones de comportamiento y el impacto social ante un hecho como el de Luis Rubiales y Jenni Hermoso.

II. Objetivos del trabajo

En el marco de este Trabajo de Fin de Grado, me propongo alcanzar varios objetivos fundamentales. En primer lugar, mi meta es proporcionar una explicación completa de los fundamentos del *marketing* digital, abordando los tipos y una breve revisión histórica. En este proceso, destacaré las diferencias entre los medios de comunicación y publicidad convencionales y las innovadoras alternativas digitales. Además, me centraré en un aspecto crítico del *marketing* digital: el *social listening*. Abordaré en profundidad su origen, utilidades y aplicaciones. Finalmente, desarrollaré un procedimiento metodológico detallado que se aplicará a un caso de estudio concreto, en este caso, la figura de Luis Rubiales. Mi objetivo es recopilar información de Twitter, la plataforma de expresión de opiniones por excelencia, y analizarla en busca de patrones y conclusiones que nos permitan comprender cómo se configura y modifica la opinión pública a lo largo del tiempo. Este enfoque integral enriquecerá nuestra comprensión de la dinámica de la comunicación en el entorno digital y su impacto en la percepción pública.

III. Marketing y publicidad en la era digital

A. ¿Qué es el *marketing* digital?

El *marketing* digital, *marketing* electrónico o *e-marketing* incluye el uso de herramientas tecnológicas para la elaboración de estrategias de mercado, comercialización, comunicación de productos/servicios, posicionamiento de la marca, segmentación de mercados, y un largo etcétera de objetivos publicitarios. Todo lo que antes se hacía a través de los medios publicitarios convencionales (prensa, radio, televisión, cine y publicidad exterior), ahora se lleva a cabo en lo que Philip Kotler (2011) denominaba el “ciberespacio”. Con este término, el economista y publicista estadounidense se refería al nuevo campo de trabajo de las marcas, donde desarrollan lo que también se ha conceptualizado como *Marketing* 4.0. o H2H (*Human-to-Human*): aquel que logra sus objetivos estratégicos a través de la interconexión de todos los participantes del mercado - B2B, B2C, C2B y C2C – e interacción directa, e incluso, personal.

Los nuevos medios de comunicación se caracterizan por tres elementos (Cely Álvarez, 1999): su presencia virtual, esto es, en un espacio “no existente” que ha sido referido previamente como ciberespacio, compuesto por imágenes y texto al cual solo se tiene acceso a través de dispositivos electrónicos; la instantaneidad de interacción entre todos los usuarios y, por último, la posibilidad intercambiar los roles de emisor y receptor, es decir,

Cuando las estrategias de *marketing* en las redes sociales dan sus frutos, es decir, consiguen su objetivo de generar una transacción, se habla de comercio electrónico (el famoso *e-commerce*).

B. ATL versus BTL

Se observa en la disminución de la circulación de periódicos impresos, en la decadencia de la televisión tradicional entre jóvenes y no tan jóvenes, en la popularidad de las plataformas de escucha en *streaming* y en el consecuente desuso de la radio; la publicidad centrada y personalizada en el cliente ha llegado para quedarse. Es lo que se ha denominado BTL (*Below-the-Line*) haciendo referencia a los anuncios dirigidos de forma enfocada a un cliente y diferenciada con respecto de otro, en contraposición con la publicidad tradicional dirigida al público masivo como un todo (ATL: *Above-the-Line*).

El origen de estos dos términos se remonta a 1954, cuando un gerente de publicidad de Procter & Gamble desarrolló un presupuesto de una campaña publicitaria donde se consideraron, en un inicio, únicamente los costes que acarrearían anunciarse en medios tradicionales, concretamente, en la televisión, en la radio, en medios impresos, en espacios exteriores y en obras de teatro. Posteriormente, cuando empleados de la compañía de suministro de bienes de consumo puntualizaron gastos que no se habían tenido en cuenta, tales como muestras gratuitas, patrocinios y descuentos customizados, se trazó una línea horizontal – literalmente una línea - bajo la cual se incluyeron aquellos.

Sobre la línea se plasmaron los gastos que supondrían los medios de comunicación convencionales. Entre ellos: la televisión, que a través de anuncios con imágenes en movimiento logran captar mayor atención de la que lo haría una imagen estática o un simple audio; a cambio, el coste por un anuncio de 20 segundos puede alcanzar los 25.000 euros en hora punta durante un programa de máxima audiencia. Por su parte, la radio, que es considerada el medio más versátil por su accesibilidad desde múltiples dispositivos, supone para el anunciante tarifas más razonables que las de otros canales de distribución

masiva. La prensa impresa ha sufrido el más notable de los decrementos en ventas. No obstante, su versión digital sigue aportando las ventajas que ofrecían en papel: una audiencia, en el caso de las revistas, o nicho geográfico, en el caso de los periódicos, específico y activamente participativo. A partir de 2001, Internet se ha convertido en el medio predominante, la red informática de búsqueda de cualquier tipo de información deseable y, por ende, un lugar de confianza al que acudir para resolver cualquier cuestión. Algunos de estos medios como la televisión o la radio, aunque más restringida, sí que llevan a cabo cierta segmentación de mercado, pudiendo jugar con el horario de emisión de los anuncios (un juguete no se mostrará a la 1 de la madrugada) o el momento en que se insertan (anunciarán una Thermomix en mitad de Masterchef).

Bajo la línea, se tuvieron en cuenta gastos igualmente publicitarios que, por aquella época, no incluían redes sociales sino eventos, concursos, sorteos, ferias, promociones y relaciones públicas. Actualmente, aunque no se ciñe a las redes sociales, las más veces son difundidos a través de estas.

El Estudio General de Medios que se muestra *infra* revela una caída del consumo en todos los medios tradicionales, situación que se refleja, por ende, en la inversión publicitaria que reciben.

Fig. 1. Inversión publicitaria en los medios tradicionales



Fuente: Asociación para la Investigación de Medios de Comunicación (2024)

Los porcentajes de la imagen hacen referencia a la audiencia/usuarios por día en todos menos en exterior y en cine que se refiere a soportes vistos/espectadores a la semana.

C. BTL, el favorito de las empresas

A pesar de que la elección entre una u otra depende en gran medida de los objetivos de *marketing* de una empresa o una campaña publicitaria en particular, y el intento del ATL de virar su estrategia a nichos cada vez más segmentados, esta publicidad encuentra su límite en el canal de proyección. Mientras que las redes sociales funcionan con algoritmos y el estudio de cada uno de los usuarios que las conforman, la publicidad a través de medios de comunicación masivos se emite a un público general e indiferenciado, más o menos amplio, pero limitadamente identificable. Consecuentemente, las acciones BTL son capaces de brindar retroalimentación precisa en menor tiempo, actuando mediante el factor sorpresa, el acercamiento, la expectativa, la interacción y el recuerdo. Además, juegan con la ventaja de poder involucrar a los consumidores de manera activa, lo que tiende a generar un mayor compromiso y conexión con la marca. La publicidad ATL, no obstante, sigue siendo elemental para determinadas marcas conocidas mundialmente, como pueden ser Coca-Cola, Procter & Gamble o Mercedes, que buscan mantener y reforzar su imagen de marca.

IV. Introducción a la inteligencia de datos

A. Concepto

El mundo se ha inmerso tan rápido en una era en la que no se puede negar que la tecnología digital es una fuerza motriz que ha estimulado el mundo de los negocios. Por tal motivo, es natural asumir que el panorama digital está impulsado, moldeado e influenciado por algo que hasta hace unos años parecía irrelevante: los datos. La forma en la que las empresas valoran los datos ha virado, se ha vuelto un ingrediente enriquecedor para casi todas las empresas. Pero verdaderamente, una vasta colección de datos, crudos y sin analizar, es precisamente eso, una vasta colección de datos. No proporcionan conocimiento, ni ayudan a las empresas a tomar decisiones, ni brindan orientación alguna. Esto es lo que se conoce como *big data*, caracterizada no solo por el

inmenso volumen de datos, pero por la velocidad, variedad y valor de los mismos (las 4 Vs del *big data*²). Y aquí es donde entra en juego la noción de inteligencia de datos.

Mayormente conocida como *data intelligence*, se trata de la disciplina que se encarga de la optimización del proceso de la recolección y gestión de los datos, a través de herramientas de inteligencia artificial, con el objeto de analizar *datasets*³ masivos y transformarlos en conocimiento. Son procesos que integran datos de diversas fuentes, aplican herramientas y técnicas analíticas avanzadas y presentan los resultados de forma fácilmente comprensible por los tomadores de decisiones.

B. *Data intelligence* en relación con *data mining* y *machine learning*

Probablemente surja confusión cuando se usan estos términos en una misma frase, por lo que, aunque después se explicará en detalle el procedimiento desde la selección de los datos de trabajo hasta los diferentes tipos de conocimiento, conviene aclarar la distinción entre ellos.

El *data mining* se refiere al proceso de síntesis de datos e inferencia de correlaciones, patrones y anomalías a través de algoritmos⁴. Entre las técnicas utilizadas en el *data mining* destacan: detección de *outliers* u objetos disímiles, asociación, *clustering*, clasificación y regresión. La primera se encarga de identificar datos que se desvían significativamente del patrón general siendo poco representativos pero que son capaces de alterar los resultados y guiarnos hacia conclusiones imprecisas. La segunda de las técnicas sirve para revelar correlaciones entre variables. El *clustering* es una de las técnicas de *data mining* más aplicadas a innumerables ciencias incluido el aprendizaje automático de máquinas, que *grosso modo* divide los datos en grupos conformados por objetos con características similares. Y, por último, la clasificación y la regresión son los dos sistemas de predicción mediante los cuales algoritmos previamente entrenados con datos predicen el *output* de nueva información con que se les provea. Concretamente, un

² Para profundizar sobre el tema, véase Marr, B. (2015). *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*.

³ Datos estructurados y organizados que se recopilan y almacenan para su posterior análisis o procesamiento, y contienen información tanto numérica como textual o gráfica.

⁴ Conjunto ordenado de operaciones sistemáticas que permite hacer un cálculo y hallar la solución de un tipo de problemas.

algoritmo de clasificación predice categorías mientras que uno de regresión predice números.

El *data mining* es un paso importante en el *knowledge discovery in database* (en adelante, KDD) - en español, descubrimiento de conocimientos en bases de datos⁵. Este término se refiere a las técnicas aplicadas para extraer conocimiento de los grandes bases de datos desestructuradas⁶. Difiere del *data intelligence* en el objetivo que persigue cada uno: este se enfoca en usar el valor creado a partir de los datos de la empresa para guiar la toma de decisiones empresariales, mientras que aquel se centra en dar solución a un problema en particular. Por su parte, el aprendizaje automático (del inglés, *machine learning*), aunque se basa en los fundamentos de la minería de datos, va más allá de identificar automáticamente correlaciones, usa algoritmos que “aprenden” cuando se los alimenta con datos de entrenamiento y aplica estos conocimientos a la creación de nuevos algoritmos. Es una tecnología que entrena a una computadora con datos de muestra para mejorar su eficiencia. Es decir, los modelos son entrenados y la máquina aprende a comportarse de acuerdo con lo que le – mandan – esos modelos.

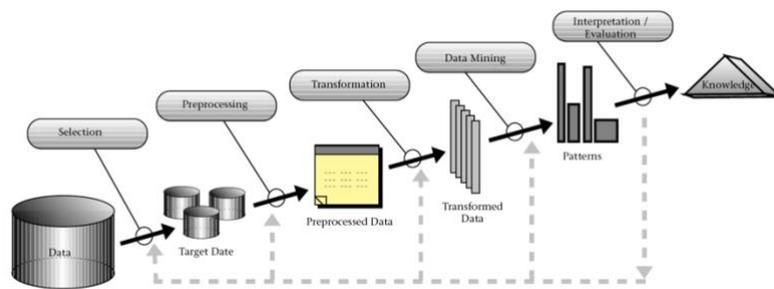
Aflora entonces una duda: ¿cómo se entrena a una maquina o cómo se obtienen patrones de tantísimos datos aparentemente inconexos? Con algoritmos. Alan Turing, matemático, científico informático y uno de los padres de la informática moderna, se encargó de formalizar este concepto como “las fórmulas matemáticas que proporcionan instrucciones a una computadora”. Son utilizados tanto en *data mining* como en *machine learning* para crear modelos, entendiendo un modelo como un conjunto de algoritmos capaz de recibir datos y generar predicciones. Según qué propósito se crea un modelo u otro. Las técnicas de *data mining* que se han explicado supra se refieren a distintos enfoques para conseguir objetivos específicos

Por tanto, el *data intelligence* incluye en su sentido el *data mining* pues son necesarias sus técnicas y algoritmos para crear modelos que puedan extraer información funcional y conocimiento de datos no procesados.

⁵ Véase Fig. 2.

⁶ Datos estructurados son los que encajan perfectamente en campos y columnas fijos en bases de datos relacionales y hojas de cálculo y los datos no estructurados aquellos que no tienen un modelo predefinido por lo que no pueden procesarse y analizarse utilizando herramientas y métodos convencionales.

Fig. 2. Proceso KDD



Fuente: Research Gate (2014)

C. Evolución histórica

Kevin Ashton fue pionero en emplear, en 1999, el término *Internet of Things* (Internet de las cosas, en adelante, IoT) para describir un sistema en el cual los objetos del mundo físico se podían conectar a Internet por medio de sensores.

A raíz de la aparición del IoT (que interconecta todos los dispositivos), y de las redes sociales (que interconectan a todas las personas), la extensa cantidad de datos heterogéneos que se producen está creciendo exponencialmente, lo que conlleva un acrecentamiento en la dificultad de manejarlos. Es el IoT el que proporciona una fuente de datos masivos que el *big data* almacena y que las herramientas de *data intelligence* transforman en información, la información en conocimiento y el conocimiento en valor.

En 1956 se escuchó por primera vez el término inteligencia artificial (en adelante, IA) en la conferencia “Dartmouth Summer Research Project on Artificial Intelligence” de John McCarthy⁷. Año tras año se fue investigando, avanzando y retrocediendo.

⁷ McCarthy, J. (1956) “Dartmouth Summer Research Project on Artificial Intelligence”, in. Hanover, New Hampshire

En 1997, la IA Deep Blue de IBM, derrotó al campeón mundial de ajedrez Gary Kasparov. Una máquina vencía por primera vez en la historia sobre la inteligencia humana más experimentada. Este acontecimiento supuso el inicio de exhaustivas inversiones en IA y en herramientas de manejo de datos por parte de muchas empresas, continuamente en aumento, y no solo relacionadas con la tecnología. Se ha descubierto el valor de saber interpretar los datos y nadie pierde su oportunidad para explotarlos.

D. La inteligencia de datos para la empresa: *business intelligence*

La primera persona en acuñar el término *business intelligence* fue Hans Peter Luhn, investigador de IBM, quien, en su artículo “A Business Intelligence System” (1958), lo definió como “la habilidad de aprehender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada”⁸. Concepto abstracto a la par que amplísimo, que supuso la base para la creación de tecnologías que permiten extraer y transformar datos para distribuirlos a través de una cadena de valor. En 1989, Howard Dresner - analista de Gartner, una compañía tecnológica de investigación - concretó la inteligencia de negocios como “conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos”. Entrada la década de los 90, comienzan a diseñarse los primeros modelos informáticos⁹, pero eran demasiado complejos para los analistas y usuarios por lo que requerían de asistencia de TI¹⁰. Actualmente, este concepto engloba muchos procesos que emplean las hipótesis extraídas del *big data* interno de la propia empresa (ventas, gastos, variación de beneficios...), para eliminar las deficiencias y ajustarse ágilmente a las fluctuaciones en el mercado, es decir, mejorar estrategias para impulsar el negocio. Análisis estadísticos, DSS (Decision Support System), visualización de datos, minería de datos, OLAP (On-Line Analytical Processing), o *dashboards* son algunas de las tecnologías del *business intelligence*, representadas en la Figura 3. No se entrará en detalle pues el tema que concierne a esta investigación no se centra en el *business intelligence* sino en el caso concreto de Luis Rubiales.

⁸ Luhn, H.P. (1958) “A Business Intelligence System”, *IBM Journal*, p. 314.

⁹ Tales como los Sistemas de Soporte a Decisiones (DSS) que permitían acceder a los datos y crear modelos y el Sistema de Información Ejecutiva (EIS), que permitía visualizarlos.

¹⁰ Departamento de tecnologías de la información.

consumidores de manera más efectiva. A través de la minería de datos multimedia (*multimedia data mining*), pueden descubrir preferencias y comportamientos que nunca antes habrían sido evidentes. Un testimonio en vídeo de un cliente satisfecho puede decir mucho más que una larga reseña o un informe aburrido.

La inteligencia de negocios utiliza datos internos de la empresa mientras que la inteligencia de medios los busca en el exterior. Se encuentran, sin embargo, en una intersección común: obtener información funcional que beneficie al negocio.

La inteligencia artificial juega un papel muy importante en cuanto al tratamiento de estos datos, tanto internos como externos. Formalmente, se refiere a “un campo de la informática que se enfoca en crear sistemas que puedan realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento y la percepción”¹¹, cualidades que han permitido una mayor inferencia en multimedia con la capacidad de percibir y razonar la información extraída de aquella y transformarla en conocimiento, en sabiduría para los humanos.

Sirva de ilustración una persona que acaba de tener una excelente experiencia gastronómica en un nuevo restaurante o que acaba de ver la peor película de su vida en el cine. Aquella, como muchas otras, comparte sus pensamientos en un blog, y comienza, además, un hilo de Twitter (actualmente, X), en el que muchos usuarios discuten acerca de sus mismas vivencias. Y es que expresar opiniones no es un comportamiento novedoso; vecinos comentando sobre sus coches nuevos, compañeros de trabajo hablando sobre las recetas que se han hecho usando la Thermomix, o una caldeada comida familiar en la que se debate acerca de si es mejor Pepsi o Coca-Cola. La diferencia es que estas opiniones se desvanecían en cuanto la conversación se daba por terminada por lo que no trascendían mucho más que de las personas involucradas. Parece que la aparentemente simple evolución de la expresión de opiniones del boca-oreja a los *comments* en redes sociales, ha creado toda una industria de datos, utilizados para incrementar la rentabilidad de innumerables negocios.

¹¹ Ministerio de Derechos Sociales y Agenda 2030 (2023) *Qué es la Inteligencia Artificial*, planderecuperacion.gob.es.

F. La Web 2.0

La Web 2.0 no es una red sino un concepto. Es la idea de que las aplicaciones o las páginas web mejoran mediante la colaboración y participación de los usuarios en ellas que consumen y mezclan datos de múltiples fuentes, al tiempo que proporcionan sus propios datos y servicios. Es lo que O'Reilly denominó "arquitectura de participación"¹². Las aplicaciones de la Web 2.0 ofrecen software como un servicio (SaaS) en constante actualización, fomentan la participación masiva y ofrecen una arquitectura (acceso, uso, herramientas) con barreras bajas y fácil accesibilidad para el usuario medio. Es decir, se considera Web 2.0 a la evolución de la World Wide Web (www) a las plataformas de diálogo e interacción directa entre los usuarios. Por esto, se llama también la Web social. En esta web, los usuarios son *prosumers*, anglicismo utilizado para denominar a los que producen contenido a la vez que consumen contenido.

V. Herramienta de escucha social: el *social listening*

A. Definición

El *social listening* es una estrategia de monitoreo y análisis de las conversaciones y las interacciones en línea que se llevan a cabo en las redes sociales y otros canales digitales. Su objetivo es comprender lo que las personas están diciendo sobre una marca, producto, industria o tema específico en las redes sociales y en la web en general.

B. El *social listening* integra el *social media management*

El *social media management* es la práctica de administrar una marca, empresa o individuo en las redes sociales. Ya que es una práctica esencial para cualquier marca que quiera desenvolverse en el entorno Web 2.0, el *social media manager* (en adelante, SMM) se encarga de diseñar la estrategia de marca, definiendo qué se hará, cuándo, cómo, por qué y para quién, habiendo una segmentación previa del público objetivo. La ejecución de dicha estrategia se llevará a cabo por otra figura a la par de importante que el SMM que es el *community manager* (en adelante, CM). A su vez, el SMM supervisará el correcto cumplimiento de los deberes de aquel. Asignar recursos financieros y distribuir el presupuesto entre las campañas.

¹² O'Reilly, T. (2005). Qué es Web 2.0. Patrones del diseño y modelos del negocio para la siguiente generación del software. *Boletín de la Sociedad de la Información de Telefónica*.

Este oficio suele centrarse en aumentar la visibilidad y la conciencia de marca o *brand awareness* para promover productos o servicios. El *social listening* es el utensilio utilizado para recopilar y estudiar los datos que proporcionan los usuarios de la web 2.0. y usarlos para beneficio propio. Concretamente, el CM es el encargado de recopilar los datos de las acciones que se llevan a cabo, pero el SMM es quien examina los informes y evalúa el resultado.

C. *Social media monitoring versus social listening*

Si bien ambos términos a menudo se utilizan indistintamente, el *social media monitoring* se centra en rastrear y recopilar datos específicos relacionados con las redes sociales, como menciones de marca, *hashtags*, publicaciones, comentarios, “me gusta” y “compartir”, es decir, realiza un seguimiento de su presencia en las redes sociales. Es más reactivo porque primero tus clientes interactúan y luego la marca responde a la pregunta “¿Qué dice la gente?”. El *social listening*, por su parte, va más allá de la monitorización, analizando conversaciones, sentimientos, patrones, y extrayendo conocimientos más significativos. De hecho, aquel es un paso esencial dentro del proceso de este (se explica en el bloque V de este trabajo). El *social listening* responde a la pregunta “¿Por qué lo dice la gente?” a partir de la cual se analizan datos y se sacan conclusiones que ayudan a tomar decisiones empresariales.

La pregunta que surge ahora es: ¿cómo captan y extraen las millones de interacciones de millones de usuarios de la red? La cuestión no se responde directamente, pues existen muchas formas de obtener dicha información, más o menos directas, más o menos complejas...plataformas que permiten programar publicaciones, realizar seguimiento de menciones y analizar datos de redes sociales como Hootsuite, Sprout Social, Buffer y Agorapulse; aplicaciones enfocadas específicamente en la escucha social y el análisis de menciones de la marca/empresa/persona en las redes sociales como Brandwatch, Talkwalker, Mention y Brand24; Mention, BrandMentions y Awario que permiten monitorear menciones de marca, palabras clave y competidores en tiempo real; o incluso las propias plataformas en su versión analítica como son Facebook Insights, Twitter Analytics, LinkedIn Analytics e Instagram Insights.

El *social listening* es un concepto sencillo que requiere de un proceso complejo en el que se profundizará en bloques posteriores.

D. Aplicaciones a la ciencia del *marketing* digital

En este bloque nos centraremos en aquellos fines exclusivos del ámbito del *marketing* digital para los cuales las empresas se sirven del *social listening*.

De entrada, la exposición de estos fines requiere una previa aclaración de los tres medios esenciales y complementarios de promoción y comunicación entre los que las marcas pueden elegir para conducir sus estrategias de *marketing* digital: *paid media*, *owned media* y *earned media*. Los primeros engloban anuncios pagados en plataformas digitales como Google Ads, Facebook Ads y espacios de publicidad de display. Estos anuncios ofrecen una forma efectiva de llegar a una audiencia específica y aumentar la visibilidad de la marca en línea. Al invertir en estos medios, las empresas pueden aumentar su alcance y generar tráfico calificado a su sitio web u otros canales digitales. Los segundos comprenden los canales de comunicación que la marca controla y gestiona directamente. Esto incluye el web, el blog corporativo, las redes sociales y las campañas de correo electrónico. Estos canales brindan a las empresas la oportunidad de transmitir su mensaje de manera coherente y construir una relación sólida y confiable con su audiencia. Los últimos se refieren a las menciones, reseñas y comparticiones generadas por los usuarios y seguidores de la marca. Estas interacciones reflejan la autenticidad y credibilidad de la marca en la mente del público. Las menciones positivas y las recomendaciones de clientes satisfechos pueden generar un efecto de bola de nieve, amplificando el reconocimiento de la marca y fortaleciendo la confianza del consumidor.

En términos de gestión de la reputación de la marca, el *social listening* posibilita la identificación y respuesta ágil a comentarios negativos, así como la monitorización de menciones positivas para optimizar el discurso, fortaleciendo así la imagen de la marca. Este análisis identifica no solo las fortalezas y debilidades en su *marketing* en redes sociales, sino también las oportunidades que podrían capitalizarse para mejorar la posición en el mercado. Para ilustrar, consideremos un periódico A que goza de una sólida reputación en lo que respecta a la credibilidad de sus noticias. El periódico opta por seguir de cerca y examinar la reputación de su marca, escuchando atentamente todo lo que se comenta sobre ella en redes. Constató que los lectores expresaban inquietud con la pérdida de autonomía del periódico. A raíz de este descubrimiento, el diario les comunicó

que se comprometía a preservar la integridad y la autonomía. Esta decisión le permitió mantener su participación en el mercado y su calificación de credibilidad.¹³

Continuando con las aplicaciones al marketing digital, el *social listening* desempeña un papel estratégico en el análisis de la competencia al proporcionar una visión detallada del entorno mercantil. En otras palabras, monitorear las campañas publicitarias, las respuestas de los usuarios, las interacciones y conversaciones de la competencia en tiempo real, otorga una perspectiva acerca de las tácticas de *marketing* más efectivas; una guía para averiguar qué resonará más con la audiencia y, sobre todo, qué es de lo que carece la competencia que demandan los usuarios (AKA consumidores). Aunado a esto, participar en las mismas discusiones que la competencia, previo estudio del contexto y del posicionamiento de los usuarios, permite posicionar la marca estratégicamente y destacar su experiencia y conocimientos. Este análisis de las mejores métricas, estrategias y resultados por parte de los mejores actores del mercado se denomina *benchmarking* y facilita la adaptación ágil a las dinámicas cambiantes del mercado.

Gracias al *social listening*, en la segmentación de mercados es posible la identificación de grupos demográficos específicos que hablan sobre la marca, lo que facilita la personalización de estrategias de marketing dirigidas a segmentos específicos. Asimismo, esta información detallada permite a las empresas crear estrategias de marketing altamente personalizadas y dirigidas a segmentos específicos de su audiencia. Al comprender mejor las necesidades y preferencias de estos grupos demográficos, las marcas pueden adaptar su mensaje, contenido y ofertas para satisfacer las demandas específicas de cada segmento, aumentando así la relevancia y efectividad de sus campañas de marketing. Además, el *social listening* no solo proporciona información estática sobre la audiencia, sino que también permite a las empresas realizar un seguimiento en tiempo real de las tendencias emergentes, cambios en la opinión pública y eventos relevantes en el mercado. Esta capacidad de monitoreo en tiempo real permite a las marcas adaptar rápidamente sus estrategias y responder de manera ágil a las necesidades y preferencias

¹³ Argintzóna, J. (2020, 4 diciembre). 12 casos de éxito aplicando social listening a la reputación de marca. *Digimind*.

cambiantes de su audiencia, maximizando así el impacto de sus esfuerzos de marketing y fomentando una mayor participación y lealtad del cliente.

Continuando con las practicas del *social listening* en el *marketing* digital, ayuda a la generación de contenido publicitario, o una presencia significativa e influyente en redes sociales. Ejemplo de ello está siendo Ryanair. Esta aerolínea ha conseguido mediante un ingenioso *copywriting*¹⁴ en redes sociales captar la atención de millones de usuarios, clientes y no clientes. El *community manager* de Ryanair se ha hecho viral por las ingeniosas respuestas a los usuarios que transmitían quejas a través *tweets* o comentarios en Instagram. Sirva de ilustración la siguiente imagen:

Fig. 4: Tweet Ryanair España



Fuente: X (2023)

La herramienta del *copywriting* es clave para darse a conocer y obtener tendencias emergentes y atractivas para el público objetivo, aumentando así el *engagement*¹⁵.

La capacidad de medir y evaluar el impacto de las acciones de marketing, tanto en línea como *offline*, es fundamental para entender la efectividad de las estrategias implementadas. El *social listening* es un magnífico aliado para lograr tal propósito. En línea, las empresas pueden usar herramientas de análisis web y redes sociales para rastrear el rendimiento de sus acciones digitales. Por ejemplo, las plataformas de análisis web

¹⁴ Proceso de producir textos persuasivos para acciones de *marketing* y ventas, como el contenido de correos electrónicos, sitios web, catálogos, anuncios y cartas comerciales.

¹⁵ Acción de generar un vínculo “emocional” entre una marca o empresa y su comunidad en redes sociales. A más interacciones con una cuenta, mayor *engagement rate* tendrá esa cuenta.

como Google Analytics permiten monitorear el tráfico del sitio web, las conversiones y el comportamiento del usuario, lo que proporciona información valiosa sobre cómo los usuarios interactúan con el contenido y las campañas en línea. Del mismo modo, las herramientas de análisis de redes sociales ofrecen métricas como alcance, impresiones, participación y menciones, que ayudan a evaluar el impacto y la recepción de las acciones en plataformas sociales.

En términos de atención al cliente, el *social listening* posibilita la resolución rápida de problemas al abordar las preocupaciones expresadas en las redes sociales, y ayuda en la identificación de problemas recurrentes. También es esencial en la identificación de *influencers*, es decir, de líderes de una temática en redes sociales capaces de mover a las masas; seleccionar campañas que causen impacto en la audiencia y, por añadidura, medir la repercusión y el retorno de inversión (ROI) de dichas campañas. Para ello, es necesario realizar el proceso que se explicará en el bloque siguiente.

E. Procedimiento

En este bloque se explicarán las fases procedimentales del *social listening* de manera impersonal, es decir, sin aplicación al caso concreto de Luis Rubiales pues será posteriormente cuando se realizará *social listening* sobre ese tema y se analizaran sus resultados.

Como primera aproximación, el *social listening* consta de cinco pasos típicos: *social media monitoring*, recopilación de datos, clasificación, procesamiento y análisis de esos datos con los consecuentes informes de investigación. Habiendo sido explicado en el bloque V, no indagaremos profundamente en el *social media monitoring*, salvando mencionar la necesaria elección de las redes sociales o canales de donde se pretenden extraer los datos (Facebook, YouTube, LinkedIn, *blogs*, *forums*, páginas web...) y las palabras clave que identifican el tema a investigar. Por ejemplo, si el fin es averiguar qué opinan los usuarios acerca de un nuevo lanzamiento, las palabras clave deberán ser tales vinculadas con la marca y con dicho producto.

Conviene, asimismo, realizar un acercamiento a la recolección de datos. Además de las redes sociales de acceso público seleccionadas previamente de las que extraeremos los datos, para acceder a redes privadas como conversaciones de WhatsApp es necesario que las personas que vayan a usar esa información privada reciban un permiso, o que

dichas personas sean participes de las conversaciones privadas. En este trabajo, no se utilizarán datos provenientes de redes privadas.

Con respecto a la clasificación, el grandísimo volumen de datos con el que se trata requiere una organización de los mismos, que con técnicas de *machine learning*

i) Identificación de objetivos y métricas

Como paso previo a cualquier proyecto, se han de identificar los objetivos que se quieren conseguir, la finalidad con la cual se realiza el trabajo. Puede variar desde conocer mejor a tu competencia o analizar la percepción de tu marca entre el público general – no solo de tu mercado objetivo – hasta estudiar el comportamiento de los usuarios de las redes en cuanto al “fenómeno Rubiales”. Para evaluar los aspectos que se desean, se necesita especificar las medidas cuantitativas o cualitativas que sirven para comprender las palabras clave. Aplicado al concreto escenario de este trabajo, siendo una de nuestras palabras clave “Rubiales”, nos centraremos en el análisis de sentimiento de los usuarios con relación al beso que le dio a Jenni Hermoso. Por ello, se evalúa

Entre las diversas métricas disponibles para realizar *social listening*, podemos resaltar¹⁶:

- Índice de presencia: seguimiento exhaustivo de menciones y referencias a tu empresa en comparación con tus competidores, lo que te permite evaluar tu posición y relevancia en el mercado. Además, te brinda la oportunidad de analizar las opiniones de tu audiencia sobre tus competidores y convertirlas a tu favor.
- Tasa de participación: la frecuencia con la que los usuarios interactúan con tu contenido al visualizarlo. Esta métrica también te permite detectar patrones estacionales, disparidades demográficas y cambios abruptos en la atención prestada a tu marca.
- Análisis de sentimiento (*sentiment analysis*): a través del análisis del lenguaje natural, puedes discernir si las menciones acerca de tu empresa son positivas, negativas o neutrales.
- Tasa de conversión: porcentaje de acciones de un usuario después de interactuar con un anuncio o cualquier otro activo digital. Es decir, el número de usuarios que

¹⁶ Cfr: Rodríguez, M., “Social Listening: Qué es y Cómo Medirlo?”, *IZO*, 19 de enero de 2022.

pasaron a ser clientes, como porcentaje del total de usuarios que visitaron el sitio web (clics / total acciones).

- Cuota de voz: análisis detallado de las menciones y alusiones a una empresa en contraste con sus competidores, lo que te permitirá evaluar tu impacto y posición en la industria. Además, te brinda la capacidad de examinar las opiniones de tu audiencia sobre tus competidores y capitalizarlas a tu favor.

En relación con las palabras clave, se decidirán atendiendo al objetivo que se persiga. En el caso concreto, buscamos términos relacionados con los sentimientos, expresiones y opiniones de los clientes sobre los acontecimientos que giraban en torno a Rubiales. Surte relevancia especificar no solo el plano geográfico en el que queremos investigar sino el idioma o idiomas en el que van a obtenerse los datos, en este caso, español, ya sea de España como de Hispanoamérica.

ii) Selección de palabras clave y plataformas

El segundo paso sería identificar las plataformas de redes sociales que se pretenden investigar y elegir las herramientas que se utilizarán. Si su público es aficionado a la tecnología y los videojuegos, puede elegir plataformas como Twitch o Mixer. Si quieres dirigirte a sitios de redes empresariales, quizá te convenga centrarte en Facebook o LinkedIn, por ejemplo. En este TFG extraeremos información de las cuatro principales redes sociales del momento: TikTok, Instagram, Twitter y Facebook, donde la mayor parte del contenido son opiniones de usuarios individuales.

Fundamentalmente, las plataformas¹⁷ de *social listening* se conectan con una *Application Programming Interface* (en adelante API) a redes sociales públicas. Una API, en español interfaz de programación de aplicaciones, es un mecanismo que permite a dos *softwares* comunicarse entre sí mediante un conjunto de definiciones y protocolos. Por ejemplo, una aplicación de mapas en tu teléfono utiliza una API de servicios de mapas, como Google Maps. Cuando introduces tu destino y solicitas direcciones, la aplicación envía esa solicitud a la API de servicios de mapas. La API procesa tu solicitud, calcula la

¹⁷ Herramientas que se basan en la búsqueda constante e indexación de datos públicos en redes sociales y páginas web, dependiendo de los filtros de búsqueda de información que se soliciten por parte del cliente - la marca-). Existen muchas plataformas de *social listening*, cada una de las cuales proporciona distintas utilidades. Por ejemplo, Hootsuite, YouScan, BrandWatch, Brand24 o Digimind Social.

ruta más eficiente y devuelve los datos a la aplicación. La aplicación utiliza esos datos para mostrarte la ruta en el mapa y darte indicaciones paso a paso. En este caso, la API de servicios de mapas actúa como un servicio externo que proporciona información de mapas, y la aplicación se integra con esta API para ofrecerte la funcionalidad de navegación. Lo mismo ocurre con las APIs de redes sociales que serán las que usemos en nuestro código.

iii) Recopilación de datos

El término *webscraping* se refiere a la extracción de datos de sitios web¹⁸. Puede llevarse a cabo por medio de procesos automatizados, como *bots* o rastreadores web, también llamados “arañas web”, para raspar datos con el uso de un lenguaje de programación como Python o R, junto con una librería¹⁹, y se trata de la técnica más utilizadas pues puede ser aplicada a cualquier sitio web público.

Las redes sociales, a diferencia de otras páginas web, publican una API que permite a los desarrolladores acceder a datos de perfiles de usuario, publicaciones, imágenes y realizar acciones como publicar contenido, compartirlo o interactuar con la plataforma social. Además, para acceder a una API, se deben realizar solicitudes que coincidan con sus requisitos de datos especificar la URL del sitio web – de las redes sociales en este caso – que se desea analizar y los elementos específicos que se desean extraer, como texto, imágenes, enlaces o metadatos. Ahora bien, solo Twitter, Facebook (Meta) e Instagram ofrecen una API. Sin embargo, la API de Twitter, a raíz de la adquisición de la plataforma por el empresario Elon Musk, exige una suscripción de pago para raspar información de tweets no publicados por uno mismo.

Las APIs tienen estructuras similares:

- URL base: es la dirección donde se encuentra desplegada la API.
- *Endpoint*: cada función o recurso de la API tiene una URL diferente, que empieza siempre por la URL base. Cada una de estas funciones o recursos se le conoce con

¹⁸ Gonzalo, M. (2024, 12 enero). “Scraping de la web: qué es, cuándo está permitido y qué dice la ley”. *Newtral*, párr. 1.

¹⁹ Colección de código desarrollado previamente, y reutilizable del que se sirven los desarrolladores para no tener que desarrollar todo desde cero y agilizar el proceso de codificación.

el nombre de *endpoint*. La documentación oficial de la API tiene información sobre los diferentes *endpoints*.

- *API reference*: es el nombre que recibe la documentación oficial de la API. Está compuesta por la descripción de los servicios, así como por ejemplos de *requests* and *responses*. También suele contar con una consola para simular peticiones a la API.
- *Requests*: peticiones a la API. Contiene los parámetros, cabeceras, método usado, URL del *endpoint*, etc.
- Autenticación: sistema de identificación y de acceso a la API. Normalmente sirve de control de que se tiene permisos para usar la API.

Por último, existen plataformas de *web scraping* como Octoparse, ParseHub o Visual Scraper. Permiten al usuario, extraer datos sin necesidad de pasar por el tedioso proceso de acceder a través de una API – por las trabas impuestas por cada red social como método de protección de datos –. Aunque actualmente ya de por sí las políticas son en demasía restrictivas, hay que tener en cuenta las implicaciones éticas y legales de los procesos *de data minings* y respetar los términos de servicio y la privacidad de los usuarios de las plataformas de redes sociales. En Europa, la minería de datos está avalada legalmente cuando tiene una finalidad de investigación científica no comercial²⁰. Adicionalmente, como se ha mencionado en la introducción de este bloque, las páginas web deben de ser de acceso público.

iv) Preprocesamiento de datos

Alrededor del 80% del total de los datos se consideran no estructurados²¹, lo que significa que no están formateados en un arreglo tabular típico de filas y columnas por lo

²⁰ Cfr: artículo 67 Real Decreto-ley 24/2021, de 2 de noviembre, de transposición de directivas de la Unión Europea en las materias de bonos garantizados, distribución transfronteriza de organismos de inversión colectiva, datos abiertos y reutilización de la información del sector público, ejercicio de derechos de autor y derechos afines aplicables a determinadas transmisiones en línea y a las retransmisiones de programas de radio y televisión, exenciones temporales a determinadas importaciones y suministros, de personas consumidoras y para la promoción de vehículos de transporte por carretera limpios y energéticamente eficientes.

²¹ Cfr: Symanto Communications. (2022, 27 julio). *Minería de textos en redes sociales: ¿cómo funciona?* Symanto.com.

que hay que “transformarlas” en un conjunto de datos estructurado, es decir, con un formato entendible para un ordenador. Aquel conjunto de información incluye comentarios en redes sociales, reseñas de productos, publicaciones en foros, transcripciones de servicios de atención al cliente... Como el lenguaje humano es intrínsecamente diverso, refleja los matices, la subjetividad y la idiosincrasia de las personas que lo utilizan. Esto, sumado a la rápida evolución del lenguaje añade una capa adicional de complejidad; la comunicación intergeneracional es un foco de confusión, y particularmente en Internet. Por todo esto, han aparecido las técnicas de procesamiento del lenguaje natural (NLP) como una herramienta que permite a las máquinas entender y analizar el significado de los datos cualitativos no estructurados, que combina modelos de lingüística computacional²², *machine learning* y *deep learning*²³ y que da a los ordenadores la capacidad de comprender textos y palabras habladas de la misma manera que los seres humanos²⁴. Juntas, estas tecnologías manipulan y analizan texto y voz mediante algoritmos y modelos computacionales que explicaremos más en detalle en el bloque siguiente. Esto es, el aprendizaje automático permite a los sistemas informáticos adaptarse a los matices y cambios en el lenguaje con mayor agilidad y precisión. Y aunque los humanos todavía tienen una comprensión superior del lenguaje, esta nueva tecnología ha allanado el camino para que las organizaciones puedan aprovechar grandes volúmenes de datos textuales no estructurados a una escala sin precedentes. Esto se traduce en mejoras significativas en la toma de decisiones y los resultados comerciales, ya que las empresas pueden extraer información valiosa de fuentes de datos que anteriormente eran difíciles de analizar de manera eficiente. Las técnicas de NLP son muchas y sus aplicaciones igual de diversas, desde hacer *social listening*, hasta programar un *chatbot* (v.g. el famoso ChatGPT).

Como paso previo, hemos de elucidar tres conceptos que se irán tratando a partir de ahora con gran frecuencia: texto, ítem, documento y corpus. Un documento se refiere

²² Es la ciencia que busca comprender y desarrollar modelos del lenguaje humano utilizando computadoras y herramientas de *software*.

²³ Es una red neuronal que enseña a los ordenadores a aprender y pensar como humanos (simula el cerebro humano a través de una combinación de entradas de datos, ponderaciones y sesgos).

²⁴ Cfr. “¿Qué es el procesamiento del lenguaje natural?”, *ibm.com*.

a una observación en forma de texto (v.g. un comentario en una publicación de Instagram); el texto es un conjunto de *tokens* y todo ese conjunto es un ítem, por tanto, un documento también es un conjunto de *tokens*; y, por último, un corpus es una colección de documentos que constituye el conjunto de datos.

Para poder aplicar métodos de *machine learning* a los problemas relativos al lenguaje natural, es indispensable transformar los datos textuales en datos digitales mediante la técnica de “bag of words”²⁵. Los tres principales métodos que se explican aquí son la *Document Term Matrix* (en adelante DTM), la *Document Term Frequency Matrix* (en adelante DTF) el *Document Term Frequency-Inverse Document Frequency Matrix* (en adelante TF-IDF). En la DTM cada fila corresponde a un documento y cada columna a una variable que caracteriza a cada registro. Para construir las variables de las columnas se puede darle el valor 1 si esa palabra aparece en el documento y 0 si no aparece (“Binary Representation”) o que esa variable tome como valor la frecuencia con la que esa palabra aparece en el documento (*Term Frequency Representation*). Cada texto estará entonces representado por un vector de ocurrencias (de frecuencias). Sin embargo, presenta un gran inconveniente: algunas palabras son por naturaleza más utilizadas que otras, lo que puede llevar al modelo a resultados erróneos. Para paliarlo surge el TF-IDF, en el que se cuenta el número de veces que aparece cada *token* en el documento/corpus para cada texto, y luego se divide por el número total de ocurrencias de esos mismos *tokens* en todo el corpus. Esto es, se asigna un peso a cada palabra en función de su frecuencia de manera que lo que ayuda a capturar mejor el contenido relevante y distintivo de cada corpus. Esta técnica proporciona un vector de pesos para cada texto en lugar de un vector de ocurrencias.

Fig. 5: Fórmula Term Frequency – Inverse Document Frequency

$$w_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

²⁵ Técnica de NLP que consiste en representar un documento de texto como un conjunto (bag) de palabras, ignorando el orden y la estructura gramatical de las palabras en el texto. *Cf*: Jaiswal, M., & Tiwari, M. (2018). Bag of words (BoW) model based recommender system using collaborative filtering technique. *Computer Science Review*, vol. 28, pp. 45-54.

Fuente: Medium (2024)

$t_{f_{x,y}}$ es la frecuencia del término x en y ;

d_{fx} es el número de documentos que contienen x ;

N es el total de documentos.

En este trabajo realizaremos *text mining* en RStudio utilizando el método de la matriz TF-IDF pues ofrece una representación más precisa del contenido textual, lo que permite una mejor comparación entre textos y una identificación más efectiva de temas, tendencias o patrones en las redes sociales. Además, dado que las redes sociales contienen una gran cantidad de contenido con una variedad de temas y lenguajes, TF-IDF es flexible y adaptable al centrarse en las palabras que son más relevantes y distintivas en el texto que nos atañe.

v) Procesamiento de datos

La limpieza de datos es el paso fundamental en el procesamiento de los mismos. Su función es eliminar datos irrelevantes y ruidosos, preparar el texto para el análisis y mejorar la precisión de los resultados. Se puede reducir el texto de muchas maneras distintas:

- Estandarización del texto pasando todos los términos a minúsculas, eliminando tildes y corrigiendo errores ortográficos.
- *Tokenization*: dividir el texto en entidades significativas, por ejemplo, en palabras unitarias (*unigram*), pares de palabras (*bigram*), tres palabras (*trigram*), dependiendo de lo que interese. Cada entidad representa una columna en la TDM. Pasar de un *string* a una lista de *strings*.
- Eliminación de términos con muy baja frecuencia, caracteres no alfanuméricos y las llamadas *stop words* (palabras vacías o sin significado relevante en el contexto concreto, y normalmente de muy alta frecuencia). En el lenguaje de R existe listas preestablecidas de *stopwords* que puede el usuario modificar a su gusto pues a veces conviene eliminar unas u otras palabras que generalmente tienen significado, pero no en el caso concreto (v.g. en un corpus sobre la industria de la automoción, las palabras coche y automóvil no aportan información útil).
- *POS tagging* (part-of-speech tagging): devuelve un conjunto de pares de la forma palabra-etiqueta gramatical, basado en su definición y su contexto, de forma que

asignar algún *token* específico a cada palabra. Por ejemplo, algunas etiquetas de NLTK POS²⁶ son: DT (determinante), FW (palabra extranjera) o PNN (nombre propio)²⁷.

- Reducir palabras a su raíz a través de *stemming* (reducir las palabras a su lexema²⁸) o lematización (reducir las palabras a su lema²⁹), lo que facilita la identificación de patrones y temas comunes.

vi) Análisis de datos: *sentiment analysis*

Los datos son sometidos a procesamiento y análisis bajo la supervisión de gestión humana³⁰. Esto implica que, después de la recopilación inicial, se realiza un trabajo más detallado para entender y extraer información valiosa de los datos recabados. La intervención humana es crucial en este proceso para garantizar una interpretación precisa y una comprensión profunda de los datos, asumiendo con cautela el margen de error inherente en cualquier técnica de procesamiento de texto. Ello permite identificar patrones, tendencias y percepciones significativas que puedan influir en la toma de decisiones empresariales. Como se ha explicado en el punto i), existen diversas métricas para evaluar los datos encontrados.

El análisis de sentimiento es una técnica de NLP y que se utiliza para determinar la actitud, opinión o emoción expresada en un texto. Los algoritmos de *machine learning* evalúan la polaridad emocional del contenido de un documento o del corpus completo³¹.

²⁶ Conjunto de bibliotecas y programas para el procesamiento del lenguaje natural.

²⁷ Cfr: Bird, S., Klein, B. & Loper, E. (2009). Categorizing and Tagging Words, *Natural Language Processing with Python*, O'Reilly Media.

²⁸ Unidad mínima con significado léxico que no presenta morfemas gramaticales.

²⁹ Forma que por convenio se acepta como representante de todas las formas flexionadas de una misma palabra.

³⁰ En una empresa, es el consultor de *business intelligence* el encargado de extraer, depurar y preparar los datos y de identificar nuevos programas y metodologías de procesamiento y análisis de datos. En este trabajo, seré yo la que llevé a cabo el proceso completo hasta llegar a las conclusiones de los resultados del *social listening*.

³¹ Cfr: ISDI España (2024). *Sentiment analysis: ¿Cómo funciona?*

En el presente trabajo, habrá un documento para cada periodo de tiempo que guarda todos los tweets emitidos en ese periodo. El análisis de sentimiento se puede hacer con diferentes paquetes con palabras a las que connotaciones según diferentes criterios. Los más utilizados para el lenguaje español son: el paquete “tidytext” que permite cargar un léxico de un archivo externo que contenga palabras y su polaridad o emoción asociada, y que trabaja con texto de estructura “tidy”, por ejemplo, un *dataframe*; y el léxico NRC (traducido al español) clasifica palabras según ocho emociones (alegría, tristeza, miedo, confianza, sorpresa, disgusto, anticipación, ira) y dos sentimientos (positivo y negativo). A tal efecto, siempre se podrán utilizar otros léxicos disponibles en repositorios online teniendo en cuenta que cada uno utilizará criterios de ponderación de sentimientos y emociones diversos.

VI. “El caso Rubiales”

A. Contexto

El día 21 de agosto de 2023, Luis Rubiales, expresidente de la Real Federación Española de Fútbol y exvicepresidente de la Unión de Federaciones Europeas de Fútbol (UEFA), asistió a la final del mundial femenino de fútbol en Sídney, Australia, en la que debutó España contra Inglaterra. La selección española se hizo con la victoria y en la entrega de la copa, las jugadoras tuvieron que saludar a varias personas relevantes como el presidente de la FIFA³², la Reina Letizia y la infanta Sofía, entre los que estaba Luis Rubiales. Cuando llegó el turno de Jenni Hermoso, Rubiales la felicita por su victoria, le pregunta “¿un piquito, Jenni?”, a lo que ella responde “sí”, y le agarra la cara para darle un beso. Ese suceso que tuvo lugar en un lapso de dos segundos aún tiene repercusiones para Rubiales a día de hoy. En el autobús que trasladaba al equipo al aeropuerto, se grabó un video en el que Jenni Hermoso enseña entre risas fotos del beso, mientras decía “como Iker y Sara” y hacía junto a sus compañeras bromas burlonas acerca del acontecimiento. El “sí” de Jenni Hermoso ha sido interpretado como un consentimiento viciado por la posición de poder de Rubiales hacia su subordinada, por el contexto en que se dio (con cámaras de televisión, al lado de la reina...). A pesar de la presión del presidente del gobierno, del de la FIFA y hasta de la ONU, para que dimitiera, aseguró que no lo iba a

³² Por sus siglas Federación Internacional de Fútbol Asociación.

hacer durante la Asamblea federativa en la que sí pidió perdón por su comportamiento en el palco y por el beso a Hermoso, que calificó de “espontáneo, mutuo, eufórico y consentido”³³. Jenni Hermoso confirmó en un comunicado - probablemente a raíz de la enorme presión social que recibió en medios de comunicación y redes sociales - que se sintió “vulnerable y víctima de una agresión” y formalizó la denuncia por agresión sexual contra Rubiales. Ya inhabilitado por 90 días por la FIFA, dimitió como presidente de la RFEF y como vicepresidente de la UEFA³⁴.

Las redes sociales se inundaron de opiniones y destacaron entre ellas numerosas reivindicaciones feministas. Cuando ocurrió, las opiniones eran variopintas, pero completamente sanas. Sin embargo, dieron un giro de 180 grados cuando se difundió el video de Jenni, el cual mostraba que no estaba nada afectada por el beso. Después, se volvieron a polarizar cuando ella declaró públicamente que se había sentido violada. Los medios de comunicación y personalidades famosas iban cambiando su discurso a medida que iba saliendo a la luz información o a medida que ciertos grupos metían presión por llevar a los focos un discurso políticamente correcto. En el bloque siguiente se hará un exhaustivo examen de cómo han ido cambiando las opiniones de redes a lo largo del breve periodo de tiempo en que Luis Rubiales era el protagonista de cualquier tertulia.

B. Utilidad para la interpretación de la opinión pública

El *social listening*, como se ha venido explicando a lo largo del trabajo, consiste en obtener información a partir de datos no estructurados, ya sean texto, de audio o audiovisuales, de redes sociales. En concreto, he monitoreado las opiniones de los “twitteros”, que me han proporcionado una visión detallada de cómo diferentes segmentos de la población perciben el suceso. En este contexto, el *social listening* permite a las organizaciones ajustar sus estrategias de comunicación en respuesta a las reacciones del público. Si se detecta una fuerte tendencia negativa, las partes involucradas, como ha sido la FIFA, la RFEF o, por supuesto, el mismo Rubiales, pueden emitir declaraciones, ofrecer disculpas o tomar medidas correctivas para abordar las preocupaciones del

“El presidente de la Federación Española, Luis Rubiales, dice que no va a dimitir tras el beso a Jenni Hermoso” (2023, 25 agosto). *RFI*.

34 Martín, J.M. (28 de diciembre de 2023). El "piquito", los gestos obscenos... el terremoto que acabó con Luis Rubiales. *La razón*.

público. De este modo, no solo se obtiene una visión detallada y precisa de la polaridad de la opinión pública sobre la polémica del beso de Luis Rubiales a Jenni Hermoso, sino que también dota a las organizaciones de la capacidad de responder de manera ágil y efectiva a las dinámicas cambiantes - y tan cambiantes - de la opinión pública en Twitter.

C. Metodología

Los datos de Twitter han sido extraídos a través de una herramienta de social listening - Octoparse - que facilita la inspección de páginas web y redes sociales y la extracción de datos concretos (en este caso, del texto de tweets según tres periodos de tiempo: del 20/08/2023 al 31/08/2023, del 1/09/2023 al 31/10/2023 y del 1/11/2023 al 31/12/2023). La decisión de dividirlo de tal manera se debe a que el hecho sucedió el 20 de agosto, y los diez primeros días fueron el *peak* de emisión de opiniones, no solo en redes sino también en medios de comunicación, donde el acontecimiento acaparaba portadas de periódicos y abría telediarios. El 25 de agosto presidió una asamblea extraordinaria de la RFEF en la que declaró que el beso había sido consentido donde se viralizó su repetida frase “¡No voy a dimitir, no voy a dimitir!”, causando otra oleada de opiniones. La dimisión de Rubiales como presidente de la RFEF el 30 de octubre supuso otro hito dentro de la cronología de la polémica. Y finalizo en nochevieja pues los días de antes fueron germen de multitud de memes navideños con relación a Rubiales.

Los datos, todos en español, han sido guardados en formato *xlsx* en mi equipo y constan de tres columnas (nombre, cuenta y tweets) y de tantas filas con tweets haya recabado en ese periodo. He omitido retweets para no sesgar los datos con redundancias sin valor. Han sido preprocesados por separado – eliminando *stopwords*, números, puntuación, links y menciones y pasando a minúsculas – en los tres documentos distintos. Asimismo, dentro del proceso de limpieza, he hallado y eliminado las 40 palabras más repetidas entre los tres documentos, así acotando más la base de búsqueda. Necesariamente, han sido tokenizados para obtener un tóken por cada palabra para poder llevar a cabo el estudio. Una vez preparados para su tratamiento, he recurrido al método de la matriz *Term Frequency-Inverse Document Frequency* que calcula los pesos de cada token permitiéndome identificar los términos más importantes dentro de ese documento específico.

Con vistas a estudiar el comportamiento emocional de los usuarios, he utilizado dos métodos para el análisis de sentimiento: con el diccionario de palabras proporcionado

por el paquete *syuzhet* de R se han clasificado el conjunto de tweets de cada documento en las ocho emociones previamente mencionadas, y con un diccionario externo descargado de GitHub, un léxico traducido al español correspondiente al diccionario AFINN del paquete *tidytext*, que contiene una lista de palabras con puntuaciones asociadas que van desde -5 a +5, indicando el grado de negatividad o positividad de cada palabra. He obtenido, igualmente, la puntuación global de cada documento de manera que, teniendo en cuenta el total de términos estudiados, puede compararse con el resultado obtenido para los otros documentos.

La decisión de dividir por periodos (en adelante P1, P2 y P3) me ha permitido analizar separadamente las opiniones y contrastar diferencias *a posteriori*. Los tweets recabados han sido analizados con respecto a su polaridad positiva, negativa o neutra, así como clasificados por las categorías sentimentales que ofrece el léxico NRC (National Research Council) de R; ira, anticipación, disgusto, miedo, alegría, sorpresa, tristeza y confianza, y los sentimientos positivo o negativo. Asimismo, Me he servido de otro léxico (AFINN) traducido al español, descargado de un repositorio *online* (Github) para realizar un análisis adicional en aras a aumentar la precisión de los resultados.

VII. Conclusiones

A. Hallazgos de la investigación

Dividiré mi exposición de conclusiones en tres partes, correspondientes a cada uno de los periodos escogidos para analizar el “fenómeno Rubiales” desde el punto de vista de los tuiteros que son, de entre los usuarios de todas las redes sociales, definitivamente los que más opiniones y críticas expresan, simple y llanamente, porque Twitter es una red nacida para ello. Finalmente, realizaré una comparación de similitudes y diferencias cronológicas entre opiniones. El total de observaciones utilizadas para el estudio, entendiendo observaciones por tweets enteros, una observación por tweet, son 7187; 258 correspondientes al primer periodo, 5326 al segundo y 1603 al tercero.

Primeramente, la matriz TF-IDF, representada en una nube de palabras, me ha allanado el camino del análisis de sentimiento, permitiéndome vislumbrar que términos son los más importantes en cada periodo de tiempo para poder entender, conociendo el contexto, los temas más recurrentes.

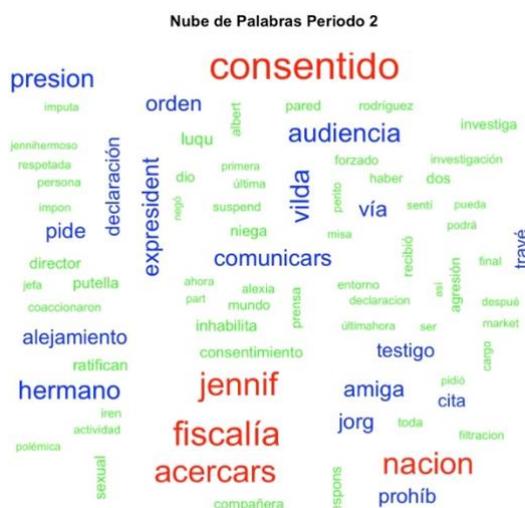
Fig. 6: Nube de palabras periodo 1



Fuente: Elaboración propia (2024)

En el P1, la palabra más repetida es “dimisión” pues tras el beso, la presión mediática, e incluso interna de la FIFA, para que Rubiales dimitiese como presidente de la RFEF fue colosal. La polémica giró en torno al presunto delito de abuso sexual, mostrándose con palabras como “besar”, “abuso”, “agresión”, “sexual” o “escándalo”. Cabezas del gobierno como Pedro Sánchez (presidente) o Irene Montero (ministra de igualdad) y todos los medios de comunicación se posicionaron en la línea feminista, defendiendo la libertad de la mujer, condenando tales actos como presuntos delitos de abuso u agresión sexual, e influenciando la opinión del público general.

Fig. 7: Nube de palabras periodo 2



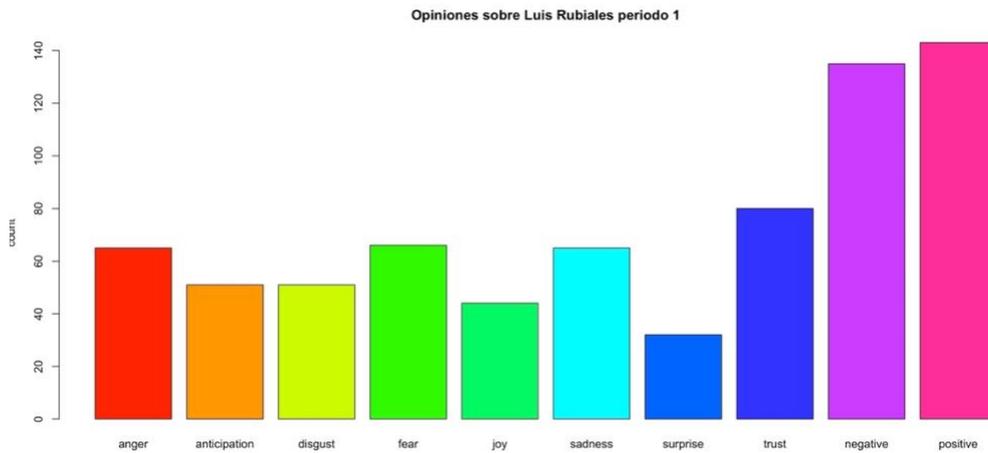
Fuente: Elaboración propia (2024)

Fig. 9: Clasificación por sentimientos del léxico NRC periodo 1

anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
65	51	51	66	44	65	32	80	135	143

Fuente: Elaboración propia (2024)

Fig. 10: Barplot de clasificación por sentimientos del léxico NRC periodo 1



Fuente: Elaboración propia (2024)

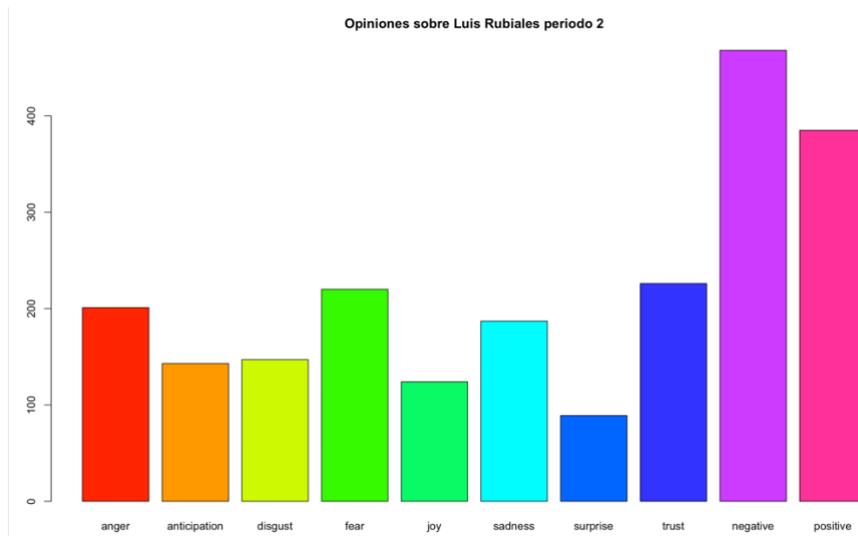
Como se ha explicado *supra*, NRC clasifica cada token en ocho emociones y dos sentimientos. Los sentimientos predominantes entre los tuiteros en los primeros 5 días tras la polémica son confianza, miedo, ira y tristeza, teniendo en cuenta que el 20 de agosto, la primera oleada de opiniones se burlaron del suceso de manera jocosa, incluida Jenni, que luego comprobamos cuando se filtró el vídeo con sus compañeras riéndose de la anécdota en el autobús. De hecho, observamos más términos positivos que negativos.

Fig. 11: Clasificación por sentimientos del léxico NRC periodo 2

anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
201	143	147	220	124	187	89	226	468	385

Fuente: Elaboración propia (2024)

Fig. 12: Barplot de clasificación por sentimientos del léxico NRC periodo 2



Fuente: Elaboración propia (2024)

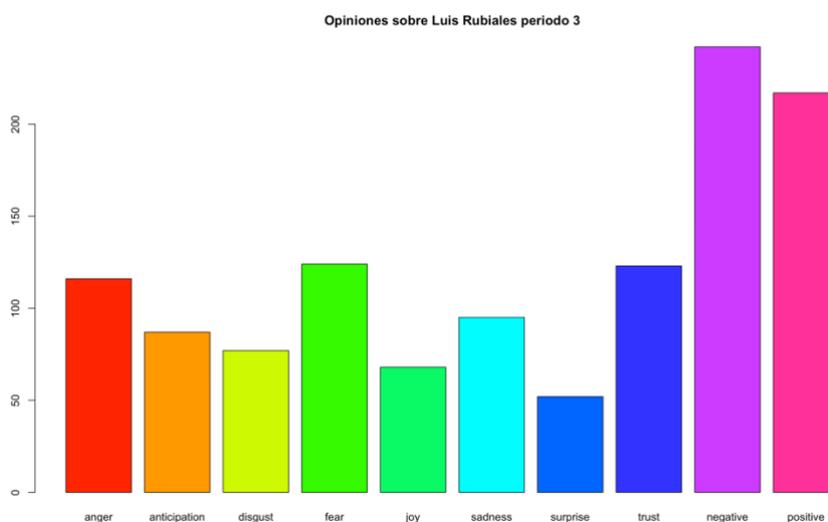
En línea con las expectativas de todos, en este segundo periodo, las críticas - y no exactamente constructivas - e insultos hacia Rubiales y hacia el género masculino en su conjunto, se dispararon vertiginosamente, impulsados por una avalancha de feminismo en todos los medios de comunicación. Líderes de gobiernos, empresas y personas de renombre se inundaron la boca de alegatos feministas. La ira sobresalta de entre los sentimientos, de la mano de la confianza, pues frente a una opinión extrema destaca siempre a la par el polo opuesto como contra opinión. El miedo y el enfado siguen en alza pues se extrapola el caso Rubiales al hipotético machismo inherente en la sociedad a raíz de la orden de alejamiento dictada por los tribunales el día 15 de septiembre.

Fig. 13: Clasificación por sentimientos del léxico NRC periodo 3

anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
116	87	77	124	68	95	52	123	242	217

Fuente: Elaboración propia (2024)

Fig. 14: Barplot de clasificación por sentimientos del léxico NRC periodo 3



Fuente: Elaboración propia (2024)

Dos meses después del suceso, los comentarios negativos siguen en alza, y si se me permite mi opinión, no creo que vaya a cambiar nunca pues Luis Rubiales no ha representado nunca una imagen especialmente agradable a ojos de la sociedad a causa de anteriores escándalos públicos, controvertidos y ordinarios. Sirva de ejemplo el obsceno gesto realizado ante la Reina Letizia de España y la infanta Sofía en el palco del estadio solo un día después del beso a Jenni Hermoso.

El análisis que se realizará a continuación no vislumbrará resultados distintos a los obtenidos con el diccionario NRC, sino que nos proporcionará una perspectiva distinta del mismo núcleo esencial.

En el P1 hay 467 palabras susceptibles de ser asignadas una puntuación, de los 3294 tokens totales del periodo 1. Esto se debe a que el léxico AFINN contiene un conjunto limitado de palabras con puntuaciones de sentimiento asignadas. Al haber realizado una unión interna, si una palabra en mi matriz de tokens no se encuentra en el léxico, no se le asignará ninguna puntuación. De haber hecho una unión externa a las palabras que no se encontrasen en el léxico se les asignaría una puntuación nula, pero hubiese devuelto una solución más sobrecargada por lo que he decidido.

La puntuación total de todo el documento de tweets del P1 es -430, clasificadas como muestra las figuras *infra*. Con una clasificación de 326 tokens con carga negativa y 138 con carga positiva, este diccionario difiere de lo mostrado con NRC. Un 37% de las palabras han sido clasificadas con una puntuación negativa de -2 como “víctima” o

“presionado”³⁵ y un 18% con -3 con “asco” o “delito”. Como contrapartida, en línea con lo expuesto anteriormente sobre el tinte positivo de muchas opiniones los primeros días de la polémica, un 25% de tweets han sido clasificados como muy positivos (≥ 2) con palabras como “ganador” o “triumfo”.

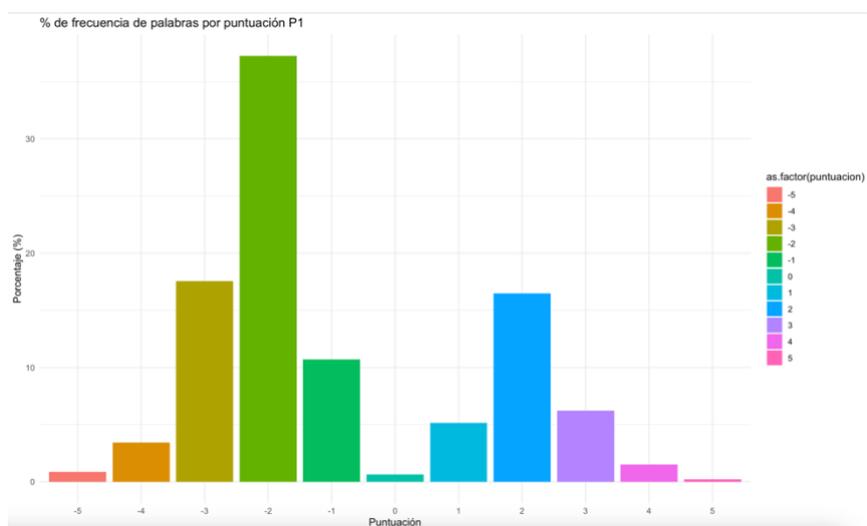
Fig. 15: Frecuencia de las palabras en base a su polaridad con AFINN P1

-5	-4	-3	-2	-1	0	1	2	3	4	5
4	16	82	174	50	3	24	77	29	7	1

Fuente: Elaboración propia

El gráfico de barras *infra* muestra los porcentajes de palabras asignadas a cada puntuación y permite una comprensión intuitiva de la distribución de las palabras, facilitando así el análisis de datos de manera visual.

Fig. 16: Barplot del porcentaje de palabras clasificadas según AFINN en el P1



Fuente: Elaboración propia (2024)

En el P2, hay 4902 palabras susceptibles de ser asignadas una puntuación, de los 49476 tokens totales del periodo 2 debido a lo previamente explicado. Con una puntuación total de -3162, la mayoría de las palabras han sido clasificadas entre el -2 y el +2. Ha de destacarse que el polo de negatividad comprende lenguaje soez, propio de Twitter.

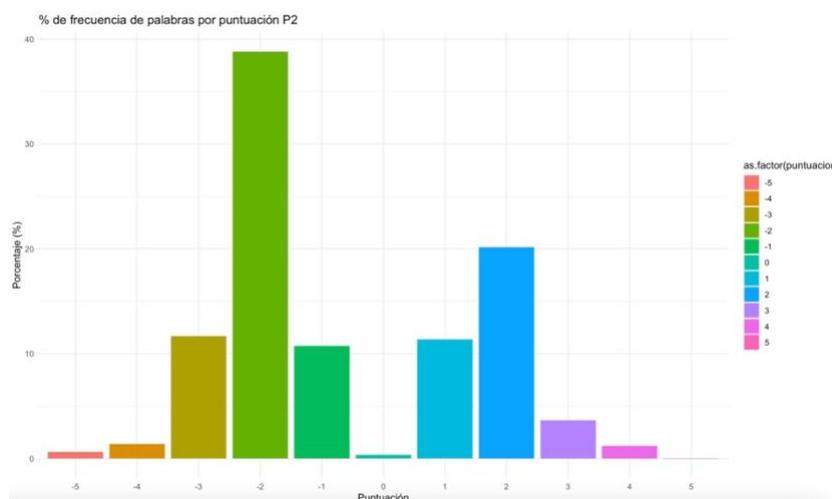
³⁵ Véase matriz s1 del código de R.

Fig. 17: Frecuencia de las palabras en base a su polaridad con AFINN P2

-5	-4	-3	-2	-1	0	1	2	3	4	5
32	68	572	1901	526	18	558	988	178	59	2

Fuente: Elaboración propia (2024)

Fig. 18: Barplot del porcentaje de palabras clasificadas según AFINN en el P2



Fuente: Elaboración propia (2024)

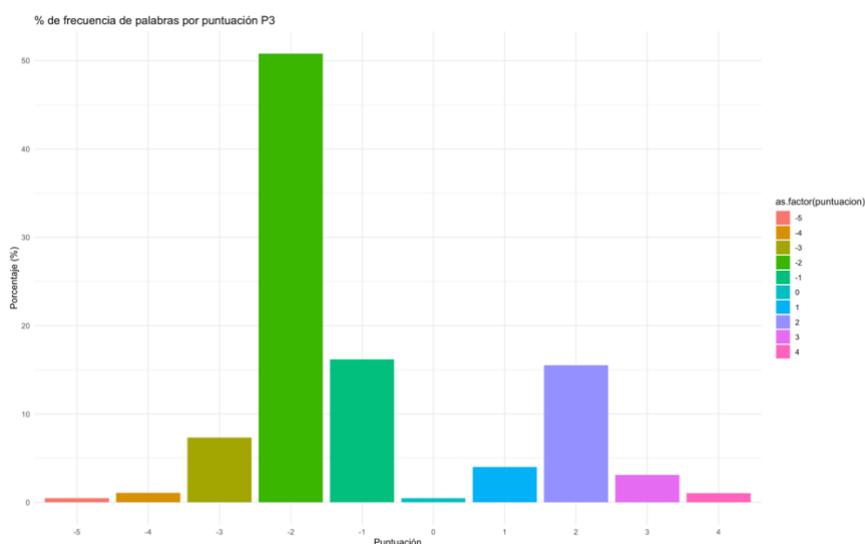
Como culmen de este estudio, llegamos al periodo 3, el cual no proporciona información muy innovadora. La puntuación predominante sigue siendo el -2, con un 50% de palabras clasificadas en él. Es una puntuación negativa reafirmante de las hipótesis iniciales de que los tweets hacia Luis Rubiales iban a tener un tinte hostil, pero sin presentar resultados excepcionalmente raros o imprevisibles.

Fig. 19: Frecuencia de las palabras en base a su polaridad con AFINN P2

-5	-4	-3	-2	-1	0	1	2	3	4
9	20	134	929	296	9	73	284	57	19

Fuente: Elaboración propia (2024)

Fig. 20: Barplot del porcentaje de palabras clasificadas según AFINN en el P3



Fuente: Elaboración propia (2024)

El tono y la naturaleza de las opiniones expresadas en plataformas digitales en torno al tema que se ha estudiado revela una predominancia de perspectivas críticas y negativamente cargadas, reflejando un clima de debate intenso en la opinión pública. Este tipo de análisis no solo proporciona una visión de las actitudes hacia los protagonistas del debate, sino que también ofrece *insights* sobre las dinámicas de interacción y percepciones dentro de la comunidad virtual que aborda este tipo de temas.

B. Futuras líneas de investigación

En la era digital y conectada en la que vivimos, el *social listening* ha emergido como una poderosa herramienta para comprender y analizar las conversaciones y opiniones expresadas en plataformas sociales y digitales. Este enfoque no solo permite a las organizaciones monitorear la percepción pública de sus marcas y productos, sino que también abre nuevas fronteras para la investigación académica en diversas áreas de interés social, política y ética. Presento la siguiente propuesta de temas en los que se puede indagar utilizando *social listening*:

- Cómo la manifestación de determinados líderes de influencia respecto al tema afecta a la opinión pública.
- Consideraciones éticas del *social listening* (ej.: protección de datos, consentimiento, confidencialidad de los clientes...).
- ¿Cómo detectar *fake news* a través del *social listening*?

Declaración de la utilización de Chat GPT u otras herramientas de IA:

Por la presente, yo, Beatriz Ibáñez Puerto, estudiante de E3 Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado “Social listening: Estudio de la reputación online de Luis Rubiales”, declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. *Brainstorming* de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 19 de junio de 2024

Firma: Beatriz Ibáñez Puerto

VIII. Bibliografía

- Alanis, E. C. G. (2022). Planeación estratégica de mercadotecnia en plataformas sociodigitales. *aDResearch ESIC International Journal of Communication Research*, vol. 28, e224. Doi: <https://doi.org/10.7263/adresic-28-224>
- Amador, O. (2023, 5 junio). *El Arte de Utilizar los Paid Media, Owned Media y Earned Media en Estrategias de Marketing Digital en México*. LinkedIn. <https://www.linkedin.com/pulse/el-arte-de-utilizar-los-paid-media-owned-y-earned-en-marketing-oscar/?originalSubdomain=es>

- Amazon Web Services, Inc. (s. f.). *¿Qué es una API? - Explicación de interfaz de programación de aplicaciones - AWS.* <https://aws.amazon.com/es/what-is/api/>
- Amazon Web Services, Inc. (s. f.). *¿Qué es el procesamiento de lenguaje natural? - Explicación del procesamiento de lenguaje natural - AWS.* <https://aws.amazon.com/es/what-is/nlp/>
- APA Style. (s. f.). <https://apastyle.apa.org>
- Argintzona, J. (2020, 4 diciembre). 12 casos de éxito aplicando social listening a la reputación de marca. *Digimind.* <https://blog.digimind.com/es/insight-driven-marketing/12-casos-exito-aplicando-social-listening-reputacion-de-marca>
- Bird, S., Klein, B. & Loper, E. (2009). Categorizing and Tagging Words, *Natural Language Processing with Python*, O'Reilly Media. <https://www.nltk.org/book/ch05.html>
- Brinkman, A. (2018, 26 diciembre). Listen and Learn: Six Tips to Master Social Media Listening. *Forbes.* <https://www.forbes.com/sites/forbescommunicationscouncil/2018/12/26/listen-and-learn-six-tips-to-master-social-media-listening/#27b7127b594c>
- Campana, N. (2023, 30 abril). ¿Qué hace un consultor BI? *Freelancermap.* <https://www.freelancermap.com/blog/es/que-hace-consultor-bi/>
- Cebotarean, E. (s. f.). Business intelligence. *Journal Of Knowledge, Management, Economics and Information Technology.* http://www.scientificpapers.org/wp-content/files/1102_Business_intelligence.pdf
- DataScientest (2023, 30 octubre). *NLP Natural Language Processing: Introducción.* DataScientest.com. <https://datascientest.com/es/nlp-introduccion>
- DataScientest (2023, octubre 30). *Inteligencia artificial: definición, historia, usos, peligros.* DataScientest.com. <https://datascientest.com/es/inteligencia-artificial-definicion#:~:text=El%20t%C3%A9rmino%20inteligencia%20artificial%20fue,la%20visi%C3%B3n%20de%20la%20IA>
- Digital Marketing tools and competition: an approach to the state of the art.* (2019). Atlantic Review of Economics, 2(3), ISSN-e 2174-3835.

- Fuchs, C. (2010). *Web 2.0, prosumption, and surveillance*. *Surveillance and Society*, 8(3), 296–301. <https://doi.org/10.24908/ss.v8i3.4165>
- Gonzalo, M. (2024, 12 enero). ‘Scraping’ de la web: qué es, cuándo está permitido y qué dice la ley. *Newtral*. <https://www.newtral.es/scraping-web-ley-mineria-ai/20230828/>
- Guerrero Carmona, F.M. & Marcos Aldón, M. (2014), “Minería de textos”, *Neodocumenta Journal of document engineering*, vol. 13. <https://www.uco.es/investiga/grupos/labindoc/images/files/neodocumenta/articulos/neodoc2014n3.pdf>
- HEAVY.AI. (s. f.). *What is Data Intelligence? Definition and FAQs*. <https://www.heavy.ai/technical-glossary/data-intelligence>
- Hicham Moad Safhi, B.F., Badr Hirchoua, B.O. y Khalil, I. (2017). “Data intelligence in the context of big data: a survey”, *Journal of Mobile Multimedia*, 13, pp. 1-4. <https://journals.riverpublishers.com/index.php/JMM/article/view/3783/2589>
- IBM (s. f.). *¿Qué es Deep Learning?* <https://www.ibm.com/es-es/topics/deep-learning>
- ISDI España (2024). *Sentiment analysis: ¿Cómo funciona?* <https://www.isdi.education/es/blog/sentiment-analysis-como-funciona#:~:text=EI%20an%C3%A1lisis%20de%20sentimiento%2C%20o,sea%20positiva%2C%20negativa%20o%20neutral.>
- Jaiswal, M., & Tiwari, M. (2018). Bag of words (BoW) model based recommender system using collaborative filtering technique. *Computer Science Review*. <https://doi.org/10.1016/j.cosrev.2018.03.001>
- Jain, A. (2024, 4 febrero). TF-IDF in NLP (Term Frequency Inverse Document Frequency). *Medium*. <https://medium.com/@abhishekjainindore24/tf-idf-in-nlp-term-frequency-inverse-document-frequency-e05b65932f1d>
- Luhn, H.P. (1958) “A Business Intelligence System”, *IBM Journal*, p. 314.
- Maher, O., Fernández, A. and Akinfenwa, D. (2023) *What are the key differences and similarities between social listening and social monitoring?* LinkedIn.

<https://www.linkedin.com/advice/1/what-key-differences-similarities-between-social#:~:text=Social%20monitoring%20is%20more%20focused,requires%20you%20to%20act%20strategically>. (acceso: 18 December 2023).

Marr, B. (2015). *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*. <http://ci.nii.ac.jp/ncid/BB1888518X>

Martín, J.M. (2023, 28 diciembre). El “piquito”, los gestos obscenos... el terremoto que acabó con Luis Rubiales. *La razón*.

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *Ai Magazine*, 27(4), 12. <https://doi.org/10.1609/aimag.v27i4.1904>

McCarthy, J. (1956). *Dartmouth Summer Research Project on Artificial Intelligence*, Conferencia de Dartmouth, Universidad Dartmouth College, Hanover, New Hampshire.

Metz, C. (2023, 3 mayo). Uno de los pioneros de la IA advierte de su peligro. *The New York Times*. <https://www.nytimes.com/es/2023/05/03/espanol/ia-peligro-google.html>

Microsoft Learn. (2023, 31 octubre). *Algoritmos de minería de datos (Analysis services - Minería de datos)*. <https://learn.microsoft.com/es-es/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=asallproducts-allversions>

Microsoft Learn. (2023, 23 diciembre). *Modelos de minería de datos (Analysis Services - Minería de datos)*. <https://learn.microsoft.com/es-es/analysis-services/data-mining/mining-models-analysis-services-data-mining?view=asallproducts-allversions>

Ministerio de Derechos Sociales y Agenda 2030 (2023) *Qué es la Inteligencia Artificial*, planderecuperacion.gob.es. [https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr#:~:text=La%20inteligencia%20artificial%20\(IA\)%20es,el%20razonamiento%20y%20la%20percepci%C3%B3n](https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr#:~:text=La%20inteligencia%20artificial%20(IA)%20es,el%20razonamiento%20y%20la%20percepci%C3%B3n).

- Montalvo, R. E. (2011). Social Media Management. *International Journal of Management & Information Systems (IJMIS)*, 15(3), 91–96. <https://doi.org/10.19030/ijmis.v15i3.4645>
- Navarro, S. (2024, 12 abril). PoS tagging con spaCy | KeepCoding Bootcamps. *KeepCoding Bootcamps*. <https://keepcoding.io/blog/como-funciona-el-pos-tagging-con-spacy/>
- O'Reilly, T. (2005). Qué es Web 2.0. Patrones del diseño y modelos del negocio para la siguiente generación del software. *Boletín de la Sociedad de la Información de Telefónica*. http://sociedadinformacion.fundacion.telefonica.com/DYC/SHI/seccion=1188&idioma=es_ES&id=2009100116300061&activo=4.do?elem=2146
- Prodware en los medios - Entrevistas, notas de prensa y artículos*. (2024, 28 mayo). Prodware | Proveedor Soluciones de Gestión Para Empresas. Partner Microsoft. [https://www.prodwaregroup.com/es-es/notas-de-prensa/#:~:text=La%20capacidad%20de%20analizar%20datos,IA\)%20se%20de%20nomina%20Data%20Intelligence.](https://www.prodwaregroup.com/es-es/notas-de-prensa/#:~:text=La%20capacidad%20de%20analizar%20datos,IA)%20se%20de%20nomina%20Data%20Intelligence.)
- Publicidad en TV*. (s. f.). oblicua.es. https://www.oblicua.es/publicidad/publicidad_en_tv.html
- Radio Francia Internacional (2023, 25 agosto). “El presidente de la Federación Española, Luis Rubiales, dice que no va a dimitir tras el beso a Jenni Hermoso”. *RFI*. <https://www.rfi.fr/es/minuto-a-minuto/20230825-el-presidente-de-la-federaci%C3%B3n-espa%C3%B1ola-luis-rubiales-dice-que-no-va-a-dimitir-tras-el-beso-a-jenni-hermoso>
- Real Academia Española. Multimedia. *Diccionario de la lengua española*. Recuperado en 10 de febrero de 2024, de <https://dle.rae.es/multimedia#>
- Rodríguez, M. (19 enero de 2022). Social listening: qué es y cómo medirlo?. *IZO*. <https://izo.es/social-listening-que-es-y-como-medirlo/#None>
- Tableau (s. f.). *¿Qué es business intelligence o inteligencia de negocios?* Tableau Software, LLC. <https://www.tableau.com/es-es/learn/articles/business-intelligence>

- Telefónica Tech (s. f.). *¿Qué son los datos no estructurados?*
<https://aiofthings.telefonicatech.com/recursos/datapedia/datos-no-estructurados>
- https://www.u-cursos.cl/ingenieria/2018/2/IN6531/1/material_docente/bajar?id=2383781
- Torres, J. I. S., & Cárdenas, E. G. (2021). *Análisis y aplicación de algoritmos de minería de datos*. Revista Perspectivas, 6(21), Pp. 71-88.
<https://doi.org/10.26620/uniminuto.perspectivas.6.21.2021.71-88>
- Tran, N. (2021). *The Role of Social Listening in a Marketing Campaign* [Tesis, Karelia University of Applied Sciences].
- Salas, E. (2018). *La línea en la publicidad. Una revisión sobre la división del ATL Y BTL*, Revista Observatorio de la Economía Latinoamericana.
<https://www.eumed.net/rev/oel/2018/08/publicidad-atl-btl.html/hdl.handle.net/20.500.11763/oel1808publicidad-atl-btl>
- Semalt.com. (s. f.). *Semalt: diferencia entre Web Scraping y Data Mining. Dos mejores herramientas para minería de datos y raspado de web | Semalt Q&A*.
<https://semalt.com/es/qa/8304-imagen-del-raspador.htm>
- Silva M, Walker J, Portillo E, Dougherty L. (2022). Strengthening the Merci Mon Héros Campaign Through Adaptive Management: Application of Social Listening Methodology. *JMIR Public Health Surveill* 2022, 8(6): e35663. Doi: 10.2196/35663
- Symanto Communications (2022, 27 julio). *Minería de textos en redes sociales: ¿cómo funciona?* Symanto.com. <https://www.symanto.com/es/blog/social-media-text-mining-how-does-it-workminería-de-textos-en-redes-sociales-como-funciona/>
- UNIR (2023, 20 noviembre). *¿Qué son las librerías en programación y para qué sirven?*
<https://www.unir.net/ingenieria/revista/librerias-programacion/>
- W. Zhu, X. Wang and W. Gao (2020) *Multimedia Intelligence: When Multimedia Meets Artificial Intelligence*, in *IEEE Transactions on Multimedia*, 22(7), pp. 1823-1835. Doi: 10.1109/TMM.2020.2969791.
- Villamar, R. M., & Salas, E. (2018). *La construcción de la marca por los medios atl y btl*. Dialnet. <https://dialnet.unirioja.es/servlet/articulo?codigo=9025075>

IX. ANEXO

Link al código de R:

https://docs.google.com/document/d/1FOq1Fw1XLsnKgD8caQmp6J_jRFLKWq22RwK6m2wo4oI/edit?usp=sharing