

13.14

DEPARTAMENTO DE COMERCIALIZACIÓN  
E INVESTIGACIÓN DE MERCADOS

# Máster Interuniversitario en Marketing y Comportamiento del Consumidor

## Trabajo Fin de Máster

**DANDO ALAS A Red Bull: Una aplicación de la Minería de Opiniones,  
para conocer qué piensan y de qué hablan los seguidores de la marca  
en Twitter.**

Autor: Filipe Steiner

Tutora: Dra. María Isabel Viedma del Jesús





## **AGRADECIMIENTOS**

---

**Un agradecimiento especial a la Profesora D. María Isabel Viedma del Jesús, al Profesor D. Antonio Gabriel López Herrera y al Profesor D. Juan Miguel Rey Pino, que dotados de paciencia y sabiduría, me han orientado en mi labor de investigación científica, con un interés y entrega singular.**

**Finalmente a Dios, a mi familia y amigos, por motivarme, dándome fuerzas para creer, que el camino puede ser difícil, pero nunca imposible de terminar.**



**“Sólo aquellos que se arriesgan a ir muy lejos, pueden llegar a saber lo lejos que pueden ir”**

**(Thomas Stearns Eliot)**



## Resumen

---

Las redes sociales, teniendo la interacción como uno de sus elementos estructurales principales, representan hoy día un importante canal de comunicación para la empresa con sus clientes, permitiendo a las empresas mantener y potenciar estrategias de marketing relacional, ya que aportan información extremadamente relevante que posibilita entender mejor a los clientes, lo que necesitan y esperan de la empresa y sus productos y/o servicios.

Con el objeto de identificar y estudiar la orientación de las opiniones expresadas por los usuarios en la red, la Minería de Opiniones y Análisis de Sentimientos surgen como un conjunto de técnicas que permiten recabar, procesar y comprender grandes cantidades de información en busca de insights y relatos que sean importantes para ambas partes.

En este contexto el objetivo principal del presente Trabajo Fin de Máster, es aplicar técnicas de Minería de Opiniones y Análisis de Sentimientos a un conjunto de publicaciones relacionadas con una marca en particular, la marca de bebidas Red Bull, y creadas por los usuarios en la red social Twitter bajo la etiqueta #givesyouwings, con el fin de reunir e interpretar datos que puedan ser utilizados por la empresa en la aplicación de estrategias de marketing relacional.

El trabajo aborda la problemática desde diferentes frentes: palabras frecuentes en lenguaje natural presentes en los tweets relacionados con #givesyouwings, hashtags más usados, usuarios más activos y con más seguidores en la red, geolocalización de la información al momento de la publicación, dispositivos usados, evaluación sentimental de la información/texto expresado en los tweets desde dos vías diferentes: subjetiva en base a encuestas y automática objetiva en base a la diferencia de los conteos de palabras de corte positivas y negativas presentes en los tweets; y comparación de resultados.

De forma general, los resultados obtenidos apuntan que los usuarios en su mayoría hablan de manera positiva, activa y diversificada en el plano internacional de Red Bull, cuando palabras como thanks, love, energy, great, enjoy, better y awesome, son las que aparecen con mayor frecuencia en sus publicaciones en Twitter.



Los temas abordados y el extenso abanico de técnicas y herramientas (lenguaje y entorno de programación R, Excel, Nvivo, Mapquest y SurveyMonkey) utilizadas en el presente trabajo, adquieren singular relevancia para las empresas en el panorama actual, donde las redes sociales representan además de un canal directo de dialogo con el cliente, un pozo repleto de valiosa información al respecto del comportamiento, perfil geográfico, económico y cultural de los mismos.

**Palabras clave:** Redes Sociales, Minería de Opiniones, Análisis de Sentimientos, Twitter, Red Bull, Self Assessment Manikin, Marketing Relacional.



# Índice

---

|   |    |
|---|----|
| 1.- Introducción .....  | 6  |
| 2.- Objetivos .....   | 20 |
| 3.- Metodología.....  | 22 |
| 3.1.- Extracción de datos .....                                       | 22 |
| 3.2.- Pre procesamiento y limpieza de los datos .....                 | 23 |
| 3.3.- Construcción de la encuesta .....                               | 24 |
| 3.4.- Análisis estadísticos realizados sobre la encuesta.....         | 25 |
| 3.5.- Explotación de los datos.....                                   | 26 |
| 3.6.- Evaluación automática-objetiva usando R .....                   | 27 |
| 3.7.- Mapa de Geolocalización .....                                   | 30 |
| 3.8.- Aspectos Éticos .....   | 30 |
| 4.- Resultados .....  | 32 |
| 4.1.- Análisis detallado de la muestra de 423 tweets.....             | 32 |
| 4.1.1.- Frecuencia de palabras y nube de frecuencia de palabras ..... | 32 |
| 4.1.2.- Etiquetas tipo hashtag (*) más frecuentes .....               | 34 |
| 4.1.3.- Usuarios con más actividad.....                               | 36 |
| 4.1.4.- Dispositivos y aplicaciones de publicación .....              | 38 |
| 4.1.5.- Mapa de geolocalización .....                                 | 40 |
| 4.1.6.- Ranking retweets.....   | 41 |
| 4.2.- Análisis detallado de la muestra aleatoria de 100 tweets .....  | 43 |
| 4.2.1.- Análisis descriptivo de la encuesta .....                     | 44 |
| 4.2.2.- Tweets de valencia positiva.....                              | 45 |
| 4.2.3.-Tweets de valencia neutro-positiva .....                       | 49 |
| 4.2.4.-Tweets de valencia neutro-negativa .....                       | 51 |
| 4.2.5.-Tweets de valencia negativa .....                              | 52 |
| 4.2.6.- Análisis de los emoticonos .....                              | 53 |
| 4.3.- Análisis automático-objetivo de los datos.....                  | 55 |
| 5.- Conclusiones .....  | 58 |
| 6.- Limitaciones .....  | 61 |
| 7.- Futuras líneas de investigación .....                             | 64 |

# 1. Introducción

---

Las nuevas tecnologías, y en especial Internet, consolidado ya como un medio y herramienta imprescindible para el marketing y la comunicación, proporcionan a las empresas un amplio abanico de posibilidades para difundir y promocionar sus marcas, productos y servicios, así como establecer relaciones con sus clientes. Resulta por tanto fundamental comprender su funcionamiento y conocer su alcance en la búsqueda de establecer una comunicación eficiente con los clientes que estén conectados a la red. Tal comunicación entre empresas y clientes debe ser comprendida y constantemente monitorizada, a fin de verificar el correcto empleo de las estrategias de marketing.

Desde su creación, la Web ha sufrido transformaciones que determinan las diferentes formas de interacción entre usuarios y la publicación de contenido. La Web 1.0, caracterizada por haber revolucionado el acceso a la información en la década de los 90, permitió que una gran cantidad de contenido fuese disponible y de libre acceso. Sin embargo permitía poca o ninguna interacción, donde el usuario mostraba un papel pasivo, sin poder interactuar, alterar o reeditar su contenido (Cormode & Krishnamurthy, 2008).



Sitios característicos de la Web 1.0 eran portales de información, programados principalmente en lenguaje HTML (sitios estáticos). En consecuencia, las herramientas típicas de Web 1.0 eran destinadas a recuperar y visualizar la información almacenada en Internet de una manera eficiente, es decir, en el menor tiempo y con el mínimo coste posible.

No obstante, en la última década, Internet ha pasado por importantes transformaciones, especialmente tras el año 2004 cuando la Web 1.0 evolucionó a la Web 2.0, término que fue acuñado por la empresa americana O'Reilly Media, para designar una segunda generación de comunidades y servicios basados en plataformas Web (Colaboradores de Wikipedia, 2014). Según O'Reilly(2007), la Web 2.0 es una plataforma que posibilita servicios y funciones online, sustituyendo la necesidad de que el usuario instale software independientes en su ordenador para utilizar determinados servicios. En este sentido, en contraste con la Web 1.0, en la era de la Web 2.0, Internet es considerado como una plataforma.



En la misma línea, para Kaplan & Haenlein (2010), ese nuevo paradigma se caracteriza por la utilización de plataformas, en las cuales, aplicaciones y contenidos pueden de manera colaborativa ser modificados constantemente por los usuarios.

En este contexto, las páginas web a su vez, dejan de ser independientes en términos de contenidos y servicios, pasando a constituir estructuras integradas múltiples, reuniendo y compartiendo con sus usuarios diversas funciones y contenidos (Berners-Lee, Fischetti, & Foreword By-Dertouzos, 2000).

Así mismo, con la Web 2.0 también hubo una mejora de los patrones de códigos, principalmente los de contenido abierto, denotándose así una mayor posibilidad de interacción de los usuarios en la red (O'Reilly, 2007).

Por lo tanto, a diferencia de las aplicaciones Web 1.0, que fueron diseñadas para crear una red de recursos de información, las aplicaciones Web 2.0 están dirigidas a crear una red humana global, la web humana, donde la información se proporciona por los individuos y no por los servidores. En la misma línea, O'Reilly (2007) resalta el concepto de inteligencia colectiva como una de las características más importantes de la Web 2.0, lo que convierte la Web en un tipo de cerebro global.



No obstante, la inteligencia colectiva en aplicaciones Web 2.0 es cuestionada por Turban, McKay, Marshall, Lee, & Viehland (2008), debido a la calidad e integridad del contenido creado por los usuarios. Según estos autores, la Web 2.0 tiene la excelente característica de aprovechar la sabiduría de muchos de sus usuarios, aunque por otro lado, también puede sufrir su ignorancia o estupidez.

Esta crítica realizada por Turban y colaboradores a la inteligencia colectiva en aplicaciones Web 2.0, se deriva de la falta de control editorial en la mayoría de la información disponible en Internet. No obstante, aunque es muy difícil establecer ese control, el riesgo de información errónea parece ser mínimo, particularmente en los sitios Web 2.0 especializados en la prestación de servicios de información (Maguregui, 2009).



Un clásico ejemplo de aplicación Web 2.0 es Wikipedia<sup>1</sup>, pues está constituida por contenido facilitado por los usuarios de la red. Un estudio publicado en la revista Nature reveló que Wikipedia es tan buena como la enciclopedia británica en cuanto a la calidad de su información (Giles, 2005). Esta conclusión se ve reforzada por el hecho de que la enciclopedia británica ha tomado la decisión de invitar al público a escribir artículos en su edición en línea, en respuesta al éxito de Wikipedia (Moore, 2009).

En definitiva, con la Web 2.0, la forma de comunicarse gana mayor participación del usuario. Lo que antes era limitado apenas a personas de alto nivel técnico, pasa a ser accesible a cualquier usuario de la red, pudiendo publicar contenidos a través de su página web, blog y redes sociales, herramientas que según Kaplan & Haenlein (2010), son definidas como aplicaciones que permiten la creación y el intercambio de información generada por los usuarios.

Tal situación proporciona la aparición de las redes sociales en internet, es decir, redes de computadoras que conectan no sólo ordenadores, sino también personas como principales exponentes de los sitios Web 2.0, que a su vez están destinados a crear una red humana. Este ambiente de colaboración y la facilidad de intercambiar información en Internet, fueron la clave esencial para la evolución de las redes sociales (Kaplan & Haenlein, 2010).



Rheingold (1996), en su libro "Comunidade Virtual" comenta que las redes sociales o comunidades virtuales no son solo lugares donde las personas se encuentran, sino también un medio para conseguir diversos fines. Esta interconexión entre ordenadores da a luz a una nueva forma de actividad colectiva, centrada en la difusión y el intercambio de información, conocimientos e intereses. Así mismo, el autor indica que las comunidades virtuales también abarcan un gran número de profesionales de diferentes áreas, resultando ser un instrumento práctico y potencial (Rheingold, 1996). Por lo tanto, una comunidad virtual, cuando está convenientemente organizada, representa una importante riqueza en términos de conocimiento distribuido, capacidad de acción y potenciación cooperativa.

---

<sup>1</sup> [http://es.wikipedia.org/wiki/Wikipedia:Acerca\\_de](http://es.wikipedia.org/wiki/Wikipedia:Acerca_de)



Lévy (2004) también defiende la participación en redes sociales como un estímulo a la formación de una inteligencia colectiva, donde se puede intercambiar información y conocimientos. Fundamentalmente, percibe el papel de las redes sociales como filtros inteligentes que ayudan a conseguir la información deseada. Para el autor, “una red de personas interesadas por el mismo tema, es más eficiente que cualquier mecanismo de búsqueda” (Lévy, 2001).

Hunt (2010), en su libro titulado “El poder de las redes sociales”, manifiesta que las redes sociales tienen un potencial especialmente poderoso para los individuos, pues transforman la manera con la que personas y grupos de profesionales se comunican entre ellos. Estos cambios en la comunicación, son el resultado de la posibilidad que ofrecen las redes sociales de transmitir y compartir información de forma colaborativa, entre personas que se encuentran físicamente distantes, a través de la creación de comunidades virtuales que tratan de temas o áreas de interés común.

En la misma línea, Tenzer, Ferro, & Palacios (2009), refieren que una red social corresponde, en su sentido más amplio, a la integración o punto de encuentro entre organizaciones, redes, asociaciones e individuos que comparten expectativas, intereses y objetivos comunes y que buscan, a través de su interacción, compartir contenidos, generar respuestas o analizar problemas de manera colaborativa.

Otros autores como Rodríguez & Santamaría (2012), definen a las redes sociales virtuales como:

“...aquellos sitios en Internet que favorecen las relaciones sociales entre los usuarios, poniendo a su disposición una gran variedad de herramientas informáticas que forman nuevas estructuras sociales o comunidades dentro del mundo virtual, conectando online a las personas, tomando como base uno o varios tipos de relaciones, como la amistad, el amor, intereses comunes, afinidades o el deseo de compartir conocimiento” (p. 237).

Además, debido al visible deseo manifestado por los usuarios de participar activamente en la producción y consumo de información, los mismos asumen el papel de productores y consumidores al mismo tiempo, acuñándose el término de “prosumers” (Toffler, 1980).

Boyd & Ellison (2007 p. 213), detallan las redes sociales como “servicios basados en la Web que permiten a los individuos (1) construir un perfil público o semi-público dentro de un sistema delimitado, (2) articular una lista de otros usuarios con los que comparten una conexión, y (3) ver su lista de conexiones y aquéllas hechas por otros dentro del sistema”.



El autor Russell (2013) dice que:

“Tenemos una necesidad profundamente arraigada de compartir nuestras ideas y experiencias, que nos da la capacidad de conectarnos con otras personas y ser escuchado, generando sentimientos de valor e importancia. Somos curiosos sobre el mundo alrededor de nosotros y en cómo organizar, manipular y utilizar la comunicación para compartir nuestras observaciones, preguntas y comprometerse con otras personas en diálogos significativos sobre nuestros dilemas” (p. 06).

El abanico de redes sociales existentes en la actualidad es muy extenso. Dentro de esta variabilidad, en este trabajo nos centraremos en el concepto de microblog<sup>2</sup> que surgió a mediados del 2006 con el objeto de permitir a sus usuarios divulgar lo que hacían en un determinado momento (Sagolla, 2009). La particularidad más elemental de un microblog, en comparación a un blog convencional, es la pequeña longitud de caracteres fijada para cada mensaje.

Twitter<sup>3</sup>, que se caracteriza como microblog, da al usuario la posibilidad de utilizar 140 caracteres para divulgar información, ideas y pensamientos en tiempo real, compartiéndola entre sus usuarios. Creado en abril de 2007, actualmente cuenta con más de 255 millones de usuarios activos mensualmente que generan más de 500 millones de mensajes al día, a los cuales se les llaman tweets (Twitter, 2014a). En ese sentido, podemos pensar en Twitter como “un servicio de mensajería de texto libre de alta velocidad y global. Una pieza glorificada de valiosa infraestructura de comunicación rápida y fácil”(Russell, 2013, p. 06).

El modelo de relaciones en Twitter permite mantenerse al día con los últimos sucesos de cualquier otro usuario, aunque ese otro usuario pueda decidir no seguirte en la misma red. Según Twitter, su misión es “proporcionar a todos los usuarios la capacidad de crear y compartir ideas e información al instante, sin ningún tipo de obstáculo” (Twitter, 2014g).

Es una red social donde un usuario puede acceder a las publicaciones creadas por otro usuario o grupo de usuarios, desde su propio muro de mensajes, lo que le permite el acceso constante al

---

<sup>2</sup> <http://es.wikipedia.org/wiki/Microblog>

<sup>3</sup> <https://twitter.com/>



contenido publicado por éstos. Es decir, el usuario 1 puede decidir "seguir" a los usuarios 2 y 3, de manera que las publicaciones hechas por 2 y 3 podrán ser visualizadas por 1, sin tener que acceder a la página personal de 2 y 3. Cada usuario puede tener una lista de personas que sigue en la red o "seguidos" (following) y de personas que le siguen o "seguidores" (followers), donde los "seguidores" tienen acceso a las publicaciones creadas por el "seguido" desde sus páginas personales (Polo, 2009).

Ante la necesidad de sus usuarios de incluir enlaces a otros medios en la web, surgen empresas como Bitly<sup>4</sup>, que ofrecen un servicio de reducción de URL, lo que proporciona al usuario mayor aprovechamiento de caracteres a la hora de transmitir su mensaje.

Otra característica interesante son las etiquetas. Están subdivididas en dos tipos: 1) de usuario, @usuario, que sirven para indicar que el mensaje va dirigido al "usuario", y 2) las hashtags, que son colocadas dentro del mensaje precedidas del símbolo almohadilla (#) y que contienen una o más palabras juntas (sin espacios), que hacen referencia al tema tratado en el mensaje. Por ejemplo, #deportederiesgo, indica que todo el contenido que esté identificado con ese hashtag, estará relacionado al tema de deportes de riesgo. Esto permite a los usuarios encontrar e intercambiar información sobre un determinado tema de interés común, además de identificar a otros usuarios a quienes también les interesa el mismo tema, corroborando la afirmación de Ampofo (2011, p. 13), de que las "redes sociales en tiempo real, son artefactos de interacción social y cultural online".



En un panorama donde uno de cada cuatro usuarios en la Web usan alguna red social (Sameh, 2013), Twitter se hace cada día más indispensable en la propagación de la información instantánea, principalmente entre usuarios y medios de comunicación. Los Trending Topics, asuntos más comentados durante el día, cambian según la llegada de nuevas noticias y están al alcance de todos los usuarios de la aplicación, proporcionando que la propia fuerza de las redes sociales se retroalimente de información y fomente los temas de mayor importancia en un determinado momento.

En definitiva, las personas confían en las redes sociales como un lugar para compartir conocimientos, experiencias, ideas, opiniones..., además de aumentar su red de contactos. Tal demanda estimula el

---

<sup>4</sup><http://bit.ly/>



florecer de comunidades virtuales relacionadas con diversos temas, reuniendo usuarios por sus gustos, ideas y negocios.

Tal panorama ha despertado el interés de las empresas que buscan una constante interacción con sus clientes. La posibilidad de intercambio de información que Internet proporciona, y en particular las redes sociales, da lugar a un valor de gran importancia en el escenario empresarial, y concretamente, en el establecimiento de relaciones con los clientes en la Web. Las empresas no tardaron mucho en identificar que las redes sociales serían una forma hábil y rápida de aplicar el marketing relacional, para estrechar relaciones con el cliente, escuchar sus necesidades y conocer en profundidad sus intereses, además de establecer un diálogo fluido y constante entre ambas partes, lo que incrementa la propuesta de valor y la fidelización de la marca (Kotler & Armstrong, 2008).

Las recomendaciones, revisiones y opiniones proporcionadas por otros usuarios que hayan experimentado algún producto o servicio, ganan cada vez más poder en las redes sociales, proporcionando un excelente entorno para estimular la práctica del eWoM (electronicword-of-mouth), comportamiento electrónico de boca-oído y sus antecedentes, como la satisfacción y lealtad del cliente (Onishi & Manchanda, 2012).

En ese ámbito, Twitter representa una propuesta de divulgación muy ágil, con mensajes directos y claros, estableciéndose como importante punto de contacto entre empresa y cliente y haciendo que las corporaciones replanteen su forma de comunicarse con los consumidores.



Los usuarios de Twitter hablan de temas que les importan, sean personales o de negocio. Esto proporciona a las empresas un excelente entorno para difundir su mensaje y conectar con sus clientes en tiempo real. Tal interacción ágil y veloz, ayuda a las empresas a mejorar su negocio (Twitter, 2014b).

Uno de los factores que hacen que las redes sociales destaquen frente a otros tipos de medios, es su capacidad de agregar y reunir información referente a sus usuarios. Éstos comparten de manera voluntaria una gran variedad de datos que en su mayoría son útiles para transacciones comerciales, obsequiando a las empresas con información, la cual recolectada y clasificada de forma adecuada, sirve como base para nuevas ideas de productos, servicios y posicionamiento en el mercado, entre otros.



Considerando este panorama, son cada vez más las empresas que utilizan las redes sociales para conseguir sus objetivos empresariales. Por ejemplo, la empresa española "Desigual" utilizó Twitter para aumentar el impacto de una de sus campañas de Año Nuevo y conseguir una mayor interacción con la audiencia, logrando incrementar su base de seguidores en la red un 762% en un periodo de 45 días, además de ampliar el alcance de sus anuncios de televisión (Twitter, 2014c).

ESPN, Entertainment and Sports Programming Network<sup>5</sup>, considerada una de las principales webs del mundo de noticias, vídeos y reportajes deportivos de actualidad, también usa Twitter para aumentar la visibilidad de su contenido en la web y fomentar las conversaciones sobre los artículos y videos que produce (Twitter, 2014d).

Otras empresas de gran porte, como Fujitsu, LG, MTV, Paramount Pictures, Porsche, Toyota y Volkswagen entre otras, usan Twitter como herramienta de comunicación para estrechar lazos con el cliente, obteniendo resultados de mucho éxito (Twitter, 2014e).

El banco ING Direct es otro ejemplo de empresa que activamente utiliza Twitter como una herramienta más de contacto con sus clientes. Dispone de un equipo de representantes para atender las consultas de los usuarios, escuchando proactivamente comentarios y quejas sobre la empresa en la plataforma y contactando con los clientes para tratar de ayudarles en lo posible (Twitter, 2014f).



No cabe duda de que redes sociales tales como Twitter, generan gran cantidad de información que resulta muy valiosa para las empresas. No obstante, esta información necesita ser filtrada y tratada con el fin de generar conocimiento y en definitiva beneficiar tanto a empresa como cliente.

Como dicen Nonaka & Takeuchi (1995, p. 159), "el intercambio de ideas, opiniones y creencias proporcionado por el diálogo, posibilita el primer y más importante paso para la creación del conocimiento: la difusión y divulgación del conocimiento táctico dentro de la red."

---

<sup>5</sup><http://es.wikipedia.org/wiki/ESPN>



En este sentido, la minería de opiniones se presenta como una herramienta eficaz capaz de recabar información y transformar esos datos en conocimiento. Chen & Zimbra (2010), la definen como una herramienta que aporta técnicas de extracción, clasificación, procesamiento y comprensión de opiniones, manifestadas en diversas fuentes de información online, comentarios en redes sociales y otros mecanismos usados por los individuos para expresar sus opiniones.

Las herramientas de minería de opiniones son ahora más que nunca necesarias para recopilar, analizar, visualizar y generar conocimiento profundo en forma de insights a partir de los conjuntos de conexiones formadas de millones de mensajes, links, posts, ediciones, actualizaciones de fotos y vídeos, revisiones y recomendaciones de productos. Estas herramientas podrían ayudar a empresas en varios sentidos:

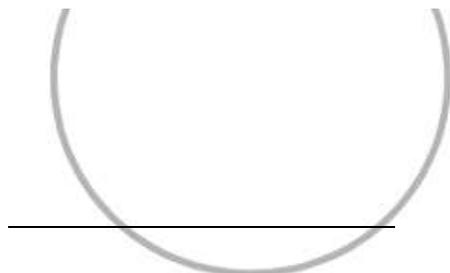
- A conocer qué se dice (ya sea bueno o no tan bueno) sobre los productos, servicios, departamentos de la empresa, o incluso de la propia empresa, en qué sentido versan las opiniones de los ya clientes o de los potenciales consumidores.
- A conocer qué están haciendo empresas competidoras en el sector.
- A conocer cuáles son los productos o servicios más relevantes en base a las opiniones positivas y buenos comentarios en las redes sociales. Así mismo, conocer cuáles son los productos o servicios menos atractivos en base a las opiniones o comentarios negativos de los usuarios.
- A detectar los posibles nichos de actuación, tanto a corto como a medio plazo, en los que la empresa pueda desmarcarse de sus competidores. Por ejemplo, un fabricante (o incluso distribuidor) de dispositivos móviles podría sondear los medios sociales y detectar que los consumidores valoran positivamente las pantallas grandes (o por el contrario negativamente las pequeñas) y en base a ello adelantarse al mercado y en consecuencia producir (o adquirir) nuevos dispositivos con estas características.



- También podrían servir para localizar que  personas o grupos son los que más activamente interaccionan con la red (tienen muchos seguidores y aportan muchas y razonadas revisiones de productos), y por tanto influyendo en su entorno. Identificar a estas personas podría ser vital para el éxito de la empresa, ya que convencerlos a ellos/as, podría significar convencer automáticamente a su entorno. Por ejemplo, la empresa podría enviarle a ellos/as los nuevos productos (o versiones previas) para que los analicen y contribuyan con opiniones, revisiones y recomendaciones positivas, que inmediatamente sus muchos seguidores van a leer.

Acciones o herramientas que trabajen en este sentido podrían encuadrarse dentro del campo de la Vigilancia Tecnológica<sup>6</sup> o la Inteligencia Competitiva<sup>7</sup>, y ayudar a personas, empresas e incluso gobiernos a tomar mejores decisiones estratégicas. La idea es sencilla, miles (o cientos de miles) de personas opinando, recomendando, o valorando positiva o negativamente algo, y en el mismo sentido, no pueden estar todas equivocadas. Una mejor comprensión de las necesidades, demandas de los usuarios, consumidores actuales o potenciales, siempre ayudara  al éxito de los productos, servicios o ideas que se quieran promover.

Herramientas de este tipo podrían incluso servir de sistemas de alerta temprana, permitiendo detectar a priori el sentir general sobre cierto tema en un marco espacio-temporal concreto. Por ejemplo, prever resultados de campañas electorales antes de que se realicen las elecciones, o prever el éxito de determinadas campañas publicitarias (Tolsan, 2012). Tales el caso de Estados Unidos que ha analizado la campaña de las elecciones presidenciales estudiando y comparando la popularidad de Barak Obama y MittRomney a través de indicadores como fans, followers, influenciadores, Tweets, Facebook posts o videos que generan más engagement, viralidad de estos contenidos y análisis de la competencia (Tolsan, 2012). Véase figura 1.



<sup>6</sup>[http://es.wikipedia.org/wiki/Vigilancia\\_tecnológica](http://es.wikipedia.org/wiki/Vigilancia_tecnológica)

<sup>7</sup>[http://www.cde.es/es/inteligencia\\_competitiva/](http://www.cde.es/es/inteligencia_competitiva/)



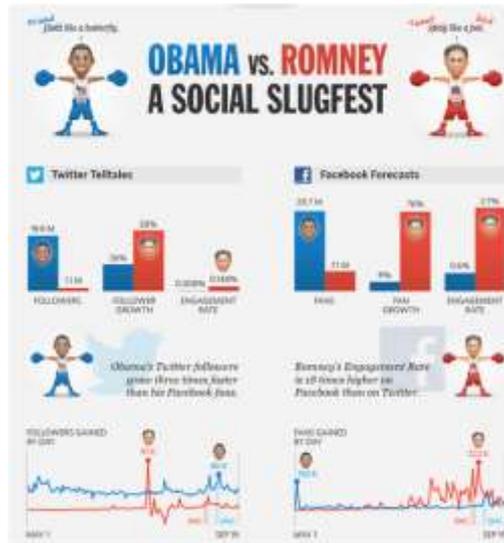


Figura 1. Análisis de las campañas electorales de Obama y Romney en las Redes Sociales.

Fuente: <https://blog.websays.com/como-detectar-la-opinion-publica-de-una-campana-electoral-gracias-a-las-redes-sociales/>

En España, también se usó la minería de opiniones durante las elecciones andaluzas para medir la influencia de los políticos en Twitter. Se centraba exclusivamente en la popularidad (nº de followers y nº de menciones) y la influencia (alcance y retweets) de la gente, generando además de un ranking, una nube de palabras para cada político (Tolsan, 2012). Véanse las figuras 2 y 3.

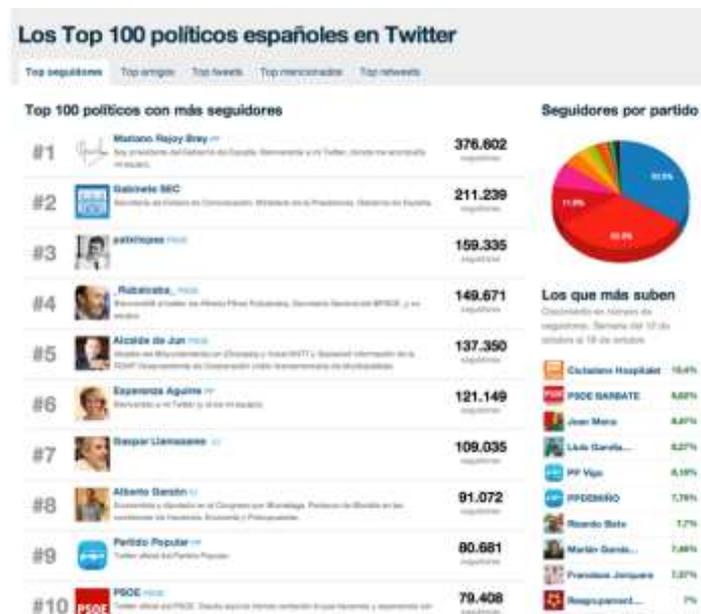


Figura 2. Ranking de políticos españoles según sus seguidores en Twitter.

Fuente: <https://blog.websays.com/como-detectar-la-opinion-publica-de-una-campana-electoral-gracias-a-las-redes-sociales/>



**Figura 3. Ranking de políticos Andaluces, según su popularidad e influencia en Twitter.**

Fuente: <https://blog.websays.com/como-detectar-la-opinion-publica-de-una-campana-electoral-gracias-a-las-redes-sociales/>

En la literatura existen propuestas que tratan de abordar este problema y surgen conceptos como Computación Afectiva, Análisis de Sentimientos, Minería de Opiniones, Análisis de Emociones y Detección de la Ironía (Ahmad, 2011). El problema no es trivial y cada uno de estos enfoques se centra en una parte concreta del análisis del lenguaje. Herramientas para la detección y cuantificación de sentimientos, opiniones, etc. existen muchas como expuesto por (Serrano-Guerrero, Olivas, Romero, & Herrera-Viedma, in litt.), pero que permitan el análisis directo de textos procedentes de medios sociales no tantas, y que permitan procesar textos de varios medios, pocas.

Por ejemplo, la Minería de Opiniones, incluyendo el Análisis de Sentimientos, tiene como objetivo extraer el contenido oculto de acuerdo con un esquema de clasificación predeterminado (Esuli & Sebastiani, 2010). Larose (2005 p. xi), la define como “el proceso de descubrir significativas y nuevas correlaciones, patrones, y tendencias por tamizado, a través de grandes cantidades de datos almacenados en los repositorios, utilizando tecnologías de reconocimiento de patrones, así como las técnicas estadísticas y matemáticas”.

Una de las vertientes del Análisis de Sentimientos es evaluar polaridad o valencia y clasificar emociones a partir de un determinado conjunto de datos, valorándose que la información sea positiva, negativa o neutra a fin de comprender lo que piensa y siente la gente (Pang & Lee, 2008). Liu (2007), entiende el análisis de sentimientos como un estudio informático de opiniones personales, expresión de emociones, evaluaciones y otros aspectos que el usuario pueda insertar en sus textos.

Se han llevado a cabo diversos estudios con el objetivo de comprender las opiniones manifestadas por los usuarios en la red social Twitter a través del Análisis de Sentimientos. Li & Li (2011),

realizaron un estudio donde se analizaron los comentarios de los usuarios en Twitter para contribuir con las empresas en cuanto a la toma de decisiones. Recurrieron a la Minería de Opiniones, a través del Análisis de Sentimientos, para hacer posible la monitorización de las opiniones de los clientes en relación a los servicios y productos de diferentes empresas, por lo que éstas podían reaccionar en el mismo momento en que sus clientes presentasen alguna queja o sugerencia.

Davidov, Tsur, & Rappoport (2010), aplicaron también Análisis de Sentimientos para clasificar emociones a partir de tweets, utilizando los mensajes de “hashtag” y los “emoticonos” contenidos en un texto. Utilizaron 50 etiquetas de Twitter y 15 emoticonos<sup>8</sup> como etiquetas de sentimiento, que les permitieron construir una clasificación para docenas de tipos de sentimiento en frases textuales cortas. Se utilizaron cuatro tipos de diferentes características (de puntuación, palabras, n-gramas y patrones) para la clasificación de sentimiento. El trabajo utiliza evaluadores humanos para comprobar que se pueden etiquetar acertadamente diferentes tipos de valencia de los tweets utilizando algoritmos automáticos.

Bello, Menéndez, Okazaki, & Camacho (2013), comparan 4 diferentes herramientas de Análisis de Sentimientos en un mismo grupo de datos obtenidos de la cuenta de IKEA en Twitter. En el mismo trabajo, exploran también las dependencias y la superposición entre los diferentes tipos de sentimientos representados por smileys y etiquetas de Twitter. Empresas como Walmart y McDonalds, entre otras, invierten en grupos de investigación especializados en Minería de Opiniones y Análisis de Sentimientos, percibiendo la gran importancia de opiniones expresadas en las redes sociales y sus implicaciones para alcanzar el éxito (Chen & Zimbra, 2010).

La empresa Campofrío utilizó la misma herramienta para medir la eficacia de su campaña publicitaria “Hazte extranjero”, analizando los tweets que contenían la palabra “Campofrío” y el hashtag #HazteExtranjero. Además, se evaluó la eficacia del Análisis de Sentimientos llevado a cabo por un software informático basado en algoritmos automáticos, en contraste con el Análisis de Sentimientos realizado manualmente por un humano. Los autores llegaron a la conclusión de que los algoritmos automáticos hacen del proceso de análisis más veloz, no obstante los resultados basados en algoritmos automáticos discrepan a los resultados conseguidos de forma manual. Se pudo verificar también que a los algoritmos automáticos les cuesta más identificar los comentarios positivos y negativos, llevando muchas respuestas a los valores neutros. Con base a esos resultados,

---

<sup>8</sup><http://es.wikipedia.org/wiki/Emoticono>



afirman que hay más robustez en el análisis de sentimiento humano (Solo Marketing, 2014). En esta línea, Pang & Lee (2008) explican que en la Minería de Opiniones, una de las principales limitaciones del Análisis de Sentimientos basada en algoritmos automáticos de textos en la Web, se relaciona con la dificultad de clasificación e interpretaciones de aspectos como la ambigüedad, ironía, sarcasmos y otras expresiones presentes en el lenguaje natural.

En cuanto al sarcasmo en Twitter, se han realizado algunos estudios con el fin de solucionar tal problema. Davidov et al. (2010), utilizan un algoritmo llamado SASI para detectar el sarcasmo en un conjunto de datos extraídos de Twitter. Reyes, Rosso, & Buscaldi (2012), en un trabajo más reciente, construyen un modelo basándose en las características textuales del lenguaje figurativo. Sus experimentos tratan de estudiar el reconocimiento del humor y de la ironía en los tweets. Muestran las dificultades encontradas al analizar tales figuras literarias, además de resaltar la importancia de considerar el lenguaje figurado para mejorar la precisión en el campo del Análisis de Sentimientos. Los ordenadores son excelentes máquinas de calcular números, sin embargo calcular sentimientos es una difícil tarea.

Teniendo en cuenta este contexto, el presente trabajo tiene como objetivo analizar y clasificar las opiniones manifestadas por seguidores de la marca Red Bull a través de tweets generados en la red social Twitter. Con este análisis se pretenden detectar los temas de interés de sus usuarios, así como identificar los sentimientos sobre los mismos a través de un análisis de la valencia o polaridad (positivo, negativo o neutral) de los comentarios, a través de algoritmos automáticos y a través de las percepciones subjetivas de personas. Los resultados obtenidos pretenden poner de manifiesto la importancia para las empresas de procesar y organizar la información presente en las redes sociales, utilizándola como base de conocimiento para entender cada vez más al detalle a sus clientes, sus comportamientos y expectativas hacia la empresa. Además, serán discutidas también implicaciones desde el punto de vista de la Minería de Opiniones y del Análisis de Sentimientos.

En las secciones siguientes se detalla de forma más concreta cuáles son los objetivos generales y específicos del trabajo, así como una descripción de la metodología seguida para la extracción, filtrado y clasificación de los tweets. Se presentarán los principales resultados obtenidos y se discutirán los hallazgos desde el punto de vista de la Minería de Opiniones, el Análisis de Sentimientos y las implicaciones para la empresa y el establecimiento de relaciones con sus clientes. Finalmente, se plantearán las limitaciones de la investigación así como las líneas futuras de actuación.



## 2. Objetivos

---

El presente trabajo considera los siguientes tres objetivos principales:

- Plantea una breve revisión sobre el estado del arte relacionado a la Minería de Opiniones y Análisis de Sentimientos y su aplicación al campo del marketing.
- Establece también utilizar las herramientas y técnicas de Minería de Opiniones con el fin de recoger, filtrar y analizar la información generada por los usuarios de Twitter bajo la etiqueta #givesyouwings, referente a la marca fabricante de bebidas energéticas Red Bull.
- Por último, establece identificar sentimientos contenidos en los tweets de los usuarios, comparando la información aportada por dos diferentes métodos de valoración: a) algoritmos automáticos y, b) valoración subjetiva a partir de una escala de valoración que va del positivo al negativo pasando por neutral-positivo y neutral-negativo.

Estos objetivos principales se concretan en los siguientes objetivos específicos:

- Detectar las palabras más usadas en términos de su frecuencia y obtener nube de palabras, lo que proporcionará la posibilidad de identificar los temas de mayor interés tratados por los usuarios.
- Detectar las etiquetas hashtags más usadas por los usuarios y generar nube de palabras con base a la frecuencia encontrada.
- Filtrar la información con el objeto de identificar y clasificar a los usuarios más activos y con mayor influencia basada en el número de seguidores.
- Identificar los dispositivos y aplicaciones más usadas por los usuarios al momento de publicar un tweet.
- Crear mapa de geo localización para identificar geográficamente el origen de la información.



- **Identificación subjetiva de la polaridad de una muestra de tweets, a través de las valoraciones de valencia.**
- **Aplicar técnicas automáticas de Análisis de Sentimientos a fin de determinar la polaridad de la muestra de tweets.**
- **Comparar las valoraciones obtenidas por el método de valoración subjetiva y el método automático.**



### 3. Metodología

---



Para la consecución y logro de los objetivos propuestos en este trabajo (ver Sección Objetivos) se han seguido los pasos (y las herramientas) que a continuación se detallan y que se resumen en la Figura 4.

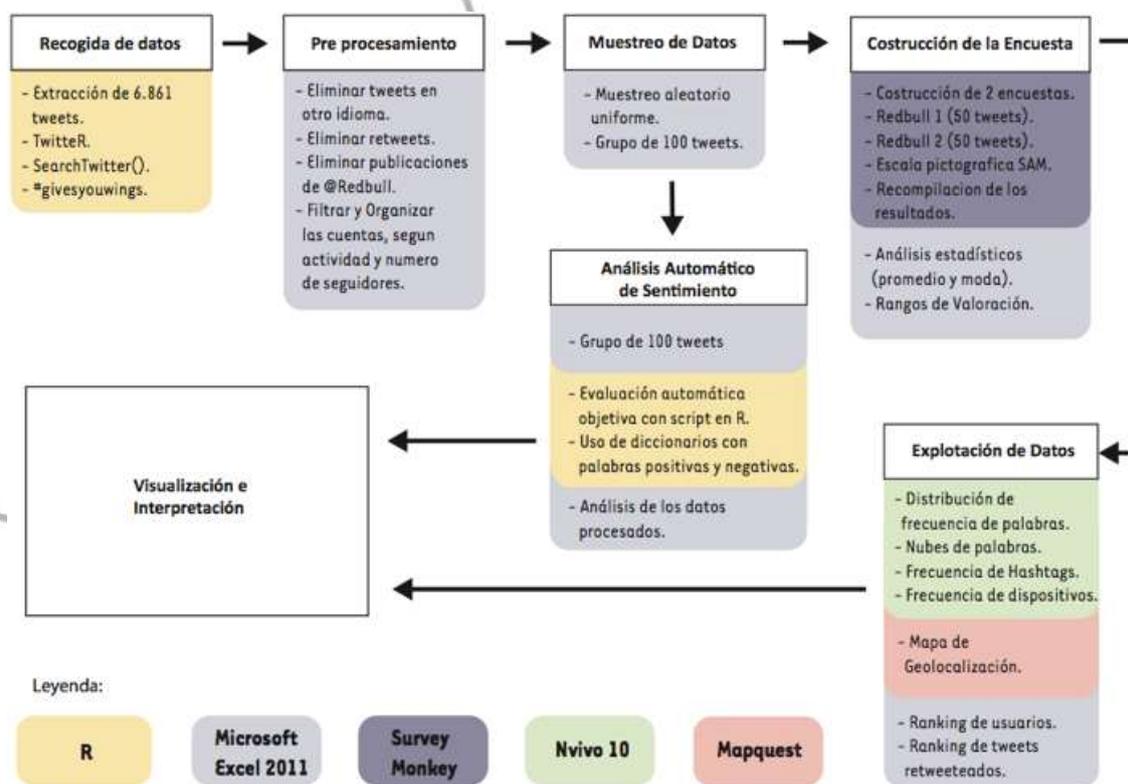


Figura 4. Flujo de trabajo y herramientas utilizadas.

Fuente: Elaboración propia.

### 3.1.- Recogida de Datos

El estudio se ha basado en la recogida de datos por internet del contenido generado por los usuarios de la red social Twitter, bajo la etiqueta [#givesyouwings](#), que está directamente relacionada a la marca Red Bull de ©Red Bull GmbH.



Para ello, inicialmente se optó por la herramienta Ncapture para NVIVO 10<sup>9</sup>. No obstante, el primer hándicap encontrado fue que, al extraer los datos, el número máximo de tweets conseguidos al día llegaba a no más de 100 publicaciones. Por tanto, para conseguir un mayor número de datos al día, se utilizó el software estadístico R para Windows<sup>10</sup>, haciendo uso del paquete *twitter*<sup>11</sup> que se trata de un entorno dotado de recursos y técnicas estadísticas que está enmarcado dentro de la plataforma GNU de licencias publicas generales (General Public License), lo que le caracteriza como software libre (R Foundation, 2002).

Los datos fueron extraídos utilizando el método *searchTwitter()* del paquete *twitter*, entre los días 03/05/2014 y 26/05/2014, mediante varios intentos. Se obtuvieron un total de 6.861 publicaciones, variando el número de información conseguida cada día. Es necesario indicar que Twitter no garantiza una recuperación de información exhaustiva, y hay siempre que asumir que los datos obtenidos son una muestra de los datos que Twitter tiene sobre un determinado tema, lo que por un lado hace que nunca se tenga garantía de haber obtenido el total de los tweets publicados, y por otro, es muy probable obtener tweets repetidos si el método *searchTwitter()* ha sido usado varias veces. Además del propio texto del tweet, *searchTwitter()* proporciona otra información de extremada importancia para nuestro estudio: número único de identificación de cada tweet en la red, nombre del usuario, fecha de publicación, cantidad de retweets generados a partir de un tweet, número de veces marcado como favorito, dispositivo utilizado para publicar el tweet en la red y su localización geográfica, cuando es permitida.



### 3.2.- Pre procesamiento y limpieza de los datos

Una vez obtenida una cantidad de datos suficientemente significativa, se pasó a la etapa de limpieza de datos. Para tal etapa del proceso, se exportaron a fichero *.CSV* los datos conseguidos con R, y éstos fueron importados a su vez a Microsoft Excel versión 2011, convirtiendo la separación por comas en columnas. El siguiente paso, consistió en eliminar publicaciones en otros idiomas

---

<sup>9</sup> [http://www.qsrinternational.com/support\\_faqs\\_detail.aspx?view=1213](http://www.qsrinternational.com/support_faqs_detail.aspx?view=1213)

<sup>10</sup> <http://www.r-project.org/>

<sup>11</sup> <http://cran.r-project.org/web/packages/twitter/>



diferentes al inglés, eliminar los tweets repetidos, eliminar los tweets generados por la marca objeto de estudio (@Redbull), además de filtrar y organizar los datos según su fecha de publicación.

Como resultado de este pre-procesamiento de datos, se consiguió un total de 2.301 publicaciones únicas, las cuales se dividen en tweets y retweets.

- **Tweets:** 532 publicaciones de las cuales, 423 fueran publicadas por cuentas de usuarios aparentemente no relacionados con la marca y 109 publicadas por cuentas relacionadas a la marca Red Bull.
- **Retweets:** Se obtuvo el número de 1.769 publicaciones de las cuales 1.747 proceden de cuentas personales aparentemente no relacionadas con la marca y 22 de cuentas relacionadas a la marca.

Para este trabajo solo se han tenido en cuenta los tweets originales y no los retweets.

### 3.3.- Construcción de la encuesta.

Para la construcción de la encuesta se utilizó la aplicación web SurveyMonkey<sup>12</sup>, para lo cual fue necesario adquirir licencia de pago por un periodo de 30 días. La divulgación se hizo vía online, por correo electrónico y a través de algunos cursos de la plataforma WebCim<sup>13</sup> del Departamento de Comercialización e Investigación de Mercados de la Universidad de Granada.

Mediante la utilización de la función [=aleatorio()] de Excel, se barajaron los 423 tweets, posibilitando obtener mediante distribución uniforme, una muestra aleatoria de 100 tweets. Para que la encuesta no fuera excesivamente extensa, ésta fue dividida en dos grupos distintos de 50 tweets cada uno, a los que llamamos Red Bull1 y Red Bull2. En el Anexo 1 se pueden ver estos dos conjuntos.

---

<sup>12</sup> <https://es.surveymonkey.com/>

<sup>13</sup> <http://marketing.ugr.es/plataforma/>



El proceso consistió en evaluar subjetivamente cada tweet en base a polaridad. Para ello se utilizó la escala pictográfica de valencia del Self Assessment Manikin, SAM, (Hodes, Cook, & Lang, 1985), que mide el grado de valencia a través de representaciones gráficas de figuras humanoides que van desde un rostro sonriente hasta un rostro serio. A través de las figuras, la escala te da un valor de 1 a 9 donde 1 significa valencia o polaridad positiva y 9 valencia o polaridad negativa. Entre estos dos extremos la escala también te permite dar puntuaciones intermedias en las figuras centrales o entre las figuras, (Bradley & Lang, 1994) (véase Figura 5).

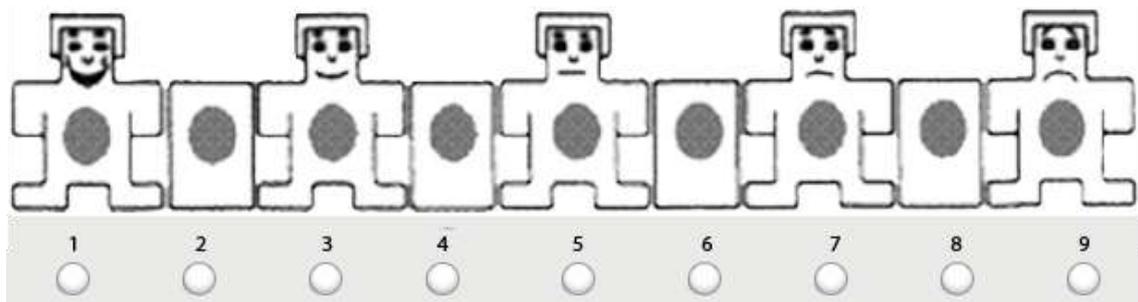


Figura 5. Escala de evaluación SAM.

Fuente: Elaboración propia a partir de (Lang, 1980).

Además de las valoraciones, se recogerá también información de variables del tipo nominal como sexo, titulación académica, conocimiento de la marca y si consume el producto, y variables ordinales como edad, nivel de estudios y volumen mensual de consumo del producto (ver Anexo 2).

cim

### 3.4.- Análisis estadísticos realizados sobre la encuesta

Para la valoración de los 100 tweets, se clasificarán los mismos según rangos de valoración (véase Figura 6).

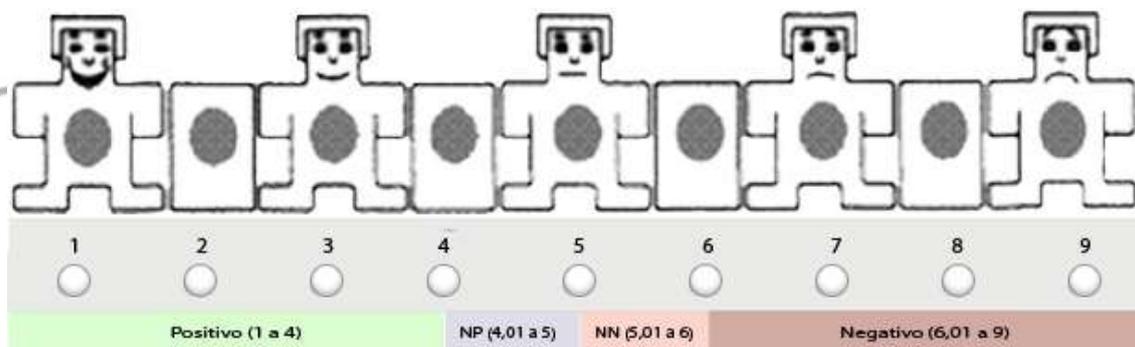
**Positivo:** comprende el intervalo de valoraciones de [1-4].

**Neutro-Positivo (NP):** comprende el intervalo de valoraciones de (4-5].

**Neutro-Negativo (NN):** comprende el intervalo de valoraciones de (5-6].

**Negativo:** comprende el intervalo de valoraciones de (6-9].





**Figura 6. Clasificación por rangos de valoración**

**Fuente: Elaboración propia a partir de (Lang, 1980).**

También se calculó la moda, obtenida desde la valoración subjetiva de cada tweet, lo que permitió identificar cuáles son las puntuaciones que más se repiten en cada caso.

### 3.5.- Explotación de los datos

En Excel se utilizan las funciones de filtrado y conteo para:

1. Identificar las cuentas con mayor actividad en #givesyouwings, según el número de publicaciones dentro del periodo estudiado.
2. Clasificar las cuentas con mayor actividad y ordenarlas en base al número de seguidores que poseían, según información conseguida manualmente en Twitter el 12/06/2014.
3. Identificar las diez cuentas no relacionadas con la marca con mayor actividad en la red y ordenadas por número de seguidores que poseían, según información conseguida manualmente en Twitter el 12/06/2014.
4. Identificar cuáles de estas cuentas son seguidas por la marca Red Bull en Twitter. Información conseguida de forma manual en Twitter el 13/06/2014.
5. Identificar las etiquetas tipo "hashtags" contenidas en la muestra.
6. Identificar los tipos de dispositivos utilizados para publicación de los tweets.

Para un análisis más en profundidad se importaron los datos desde Microsoft Excel a la herramienta de investigación cualitativa NVIVO10.0.418.0 SP4 (32bits)<sup>14</sup>, versión de prueba de 30 días, y se realizó un conteo de las palabras más frecuentes utilizando 2 grupos de datos, por un lado el grupo de 423 tweets y por otro lado la muestra aleatoria compuesta por 100 tweets. Se realizó un filtrado con coincidencias, incluyendo palabras de la raíz y de longitud mínima de 5 caracteres, de manera que, preposiciones del lenguaje natural o expresiones web como por ejemplo `http://` quedaran fuera del conteo. Además se aplicaron filtrados de palabras vacías o irrelevantes al estudio (ver Anexo 3).

Con los listados de frecuencias se construirán nubes de palabras para cada grupo de datos, haciendo posible una mejor visualización de los resultados. Con el mismo objetivo, también se construirán nubes de frecuencias de palabras con base a los listados de etiquetas tipo “hashtags” y clase de dispositivos utilizados para la publicación de los tweets en la red.

### 3.6.- Evaluación automática-objetiva usando R

Sobre el conjunto de 100 tweets evaluados subjetivamente mediante la encuesta SAM, se procede a realizar también una evaluación objetiva-automática. Diferentes métodos procedentes del campo de la inteligencia artificial o de la minería de datos pueden ser usados para este fin, como redes neuronales, métodos bayesianos, SVM (supervector machine), (Bhuta & Doshi, 2014; Hassan, Ameen, Hamed, & Mahmoud, 2013).

De entre todos ellos, para este TFM escogemos un método sencillo, pero ampliamente usado en la literatura (Hu & Liu, 2004; Liu, Hu, & Cheng, 2005), que es aquel basado en el cálculo de la diferencia del número de palabras consideradas positivas y del número de palabras consideradas negativas para cada tweet.

Es vital disponer de un conjunto amplio y validado de palabras consideradas positivas y negativas. En la tabla 1 se pueden ver algunos ejemplos.

---

<sup>14</sup> <http://www.qsrinternational.com/about-qsr.aspx>



Se decide usar el conjunto "A list of positive and negative opinion words or sentiment words for English<sup>15</sup> (unas 6800 palabras en inglés)" (Hu & Liu, 2004; Liu et al., 2005).

| Positivas   | Negativas   |
|-------------|-------------|
| abound      | abnormal    |
| abounds     | abolish     |
| abundance   | abominable  |
| abundant    | abominably  |
| accessible  | abominate   |
| accessible  | abomination |
| acclaim     | abort       |
| acclaimed   | aborted     |
| acclamation | aborts      |
| accolade    | abrade      |

Tabla 1. Ejemplos de 10 palabras positivas y 10 negativas.

Para el análisis y conteo de palabras positivas y negativas de cada tweet usamos una adaptación del script programado en R que presentamos a continuación y que fue originalmente publicado<sup>16</sup> por el Profesor Jeffrey Breen (Breen, 2012), (véase Figura 7):



<sup>15</sup> <http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar>

<sup>16</sup> <http://jeffreymbreen.wordpress.com/2011/07/04/twitter-text-mining-r-slides/>



```

library (plyr)
library (stringr)

score.sentiment = function(sentences, pos.words, neg.words, .progress='none')
{
  require(plyr)
  require(stringr)

  scores = laply(sentences, function(sentence, pos.words, neg.words) {
    # limpiamos los textos (singos de puntuación, espacios adicionales, etc.)
    sentence = gsub('[:punct:]', '', sentence)
    sentence = gsub('[:cntrl:]', '', sentence)
    sentence = gsub('\\d+', '', sentence)
    sentence = tolower(sentence)
    # y sacamos la lista de palabras de cada tweets
    word.list = str_split(sentence, '\\s+')
    words = unlist(word.list)
    # comparamos las palabras de cada tweet con los grupos de palabras de
    # positivas y negativas
    pos.matches = match(words, pos.words)
    neg.matches = match(words, neg.words)
    pos.matches = !is.na(pos.matches)
    neg.matches = !is.na(neg.matches)

    # calculamos la polaridad en base a la resta entre la cantidad de palabras
    # positivas y palabras negativas contabilizadas:
    score = sum(pos.matches) - sum(neg.matches)
    return(score)
  }, pos.words, neg.words, .progress=.progress )

  scores.df = data.frame(score=scores, text=sentences)
  return(scores.df)
}

#Cargamos las listas de palabras positivas y negativas
pos.words <- scan('positive-words.txt', what='character', comment.char=';')
neg.words <- scan('negative-words.txt', what='character', comment.char=';')

#Importamos el fichero de 100 tweets pre-procesado en Excel
Redbull <- read.csv("Redbull.csv")
Redbull$text<-as.factor(Redbull$text)

#Evaluamos los 100 tweets
Redbull.scores = score.sentiment(Redbull$text, pos.words,neg.words, .progress='text')

#Escribimos los resultados a un fichero .csv para seguir trabajando con Excel
write.csv(Redbull.scores,file="RedbullScores.csv",row.names=TRUE)

```

**Figura 7. Función usada para calcular la polaridad de un tweet.**



### 3.7.- Mapa de geolocalización

Para determinar de dónde proceden las publicaciones relacionadas con **\*givesyouwings**, se crea el mapa de geolocalización de los tweets usando las informaciones de **latitud y longitud** extraídas del grupo de 2.301 publicaciones iniciales. Se insertaron los datos manualmente en el buscador de la herramienta Mapquest<sup>17</sup> de acceso gratuito, perteneciente a © 2013 MapQuest, Inc., lo que posibilitó visualizar exactamente en qué lugar del globo estaban esos usuarios cuando realizaron determinada publicación en Twitter.

Es necesario indicar que esta información de longitud y latitud no está siempre disponible, ya que es necesario que el usuario haya dado su consentimiento para que su dispositivo y/o aplicación compartan esta información. También es necesario que el dispositivo en cuestión tenga posibilidad de geolocalizarse (por ejemplo que disponga de gps y éste esté activo). En la muestra de datos obtenida, relativamente pocos tweets estaban geolocalizados, lo que limitó nuestro análisis de esta variable.

### 3.8.- Aspectos éticos

Los aspectos éticos son tema de suma importancia en el ámbito de la investigación, sobre todo cuando se trata del mundo virtual o netnografía, término acuñado por el autor R.V. Kozinets (Kozinets, 1998). Este autor propone una serie de consideraciones referentes a los temas de tal naturaleza. Por ello, nuestra investigación sólo cuenta con información que está públicamente disponible en internet (Kozinets, 2010). Aun así, con el fin de cuidar especialmente el anonimato de las cuentas de los usuarios en Twitter, se han sustituido sus nombres por la palabra **Usuario**, cuando aparece sola, o pueden derivar según el alfabeto cuando aparezcan en grupo, por ejemplo, **Usuario A**, **Usuario B**, **Usuario C** y así sucesivamente.

Se han verificado también los términos referentes a la política de privacidad en Twitter. Ellos puntualizan que todo y cualquier usuario que utiliza esta red social, está de acuerdo que Twitter

---

<sup>17</sup> <http://www.mapquest.com>



haga pública informaciones tales como, los mensajes que “twitteas” o tweets, metadatos, fechas de los tweets, listas personales, personas a las que sigues, lista de favoritos, así como otras informaciones que se generan por el uso de los servicios en esta red. Según ellos, por defecto, casi siempre se publica la información facilitada por el usuario hasta el momento en que él mismo la elimine de la red. Twitter además, da al usuario la posibilidad de alterar determinadas configuraciones para elevar el nivel de privacidad de sus publicaciones (Twitter, 2013).

Twitter aclara que la información pública referente al perfil de usuario y tweets públicos, está disponible y puede ser utilizada por motores de búsquedas, API's, y/o empresas asociadas de Twitter. Además, recuerdan la responsabilidad que debe tener el usuario al compartir cualquier contenido en Twitter, pues se trata de una herramienta dotada de gran alcance y velocidad en la diseminación de información en la red (Twitter, 2013).



## 4. Resultados

---

Una vez expuesta la metodología llevada a cabo en el presente trabajo, en esta sección presentamos los resultados obtenidos tras su aplicación. Los resultados serán expuestos en el siguiente orden: a) análisis detallado de la muestra de 423 tweets publicados por cuentas aparentemente no relacionadas a la marca, y obtenidos tras el proceso de filtrado y limpieza sobre el total de 2301 tweets únicos totales obtenidos de Twitter y asociados al hasta #givesyouwings; b) análisis detallado del subconjunto de 100 tweets (aleatoriamente extraídos del conjunto de 423) usados para la evaluación sentimental subjetiva, así como los resultados de tal evaluación, y c) evaluación sentimental automática del conjunto de 100 tweets del punto anterior y comparación de resultados con respecto a los datos subjetivos.

### 4.1.– Análisis detallado de la muestra de 423 tweets

#### 4.1.1.– Frecuencia de palabras y nube de frecuencia de palabras

Este grupo de datos está compuesto por 423 tweets que son resultado del filtrado realizado al grupo de 2.301 tweets únicos conseguidos inicialmente de Twitter. Estos 423 tweets fueron publicados por usuarios particulares no asociados a la marca<sup>18</sup>. Sobre este conjunto de datos, se realizó un conteo de palabras en términos de frecuencia (véase Tabla 2 y para el listado completo consultar Anexo4).

Como resultados se obtuvieron:

- Las 100 palabras (junto con sus variaciones) más frecuentes con longitud mínima de 5 caracteres.
- El conteo indicador de frecuencia de cada palabra (junto con su variación).
- El porcentaje ponderado.

---

<sup>18</sup>Sin más información que la puramente obtenida vía Twitter no somos capaces de determinar si todos los usuarios que a priori se presentan como particulares en realidad lo son y no son parte de la compañía. Para el estudio llevado a cabo en este trabajo tenemos que asumir que así es. Pudiera suceder que la marca (como parte de su estrategia de social media) esté usando alias con apariencia personal con objeto de aumentar su impacto en la red, como suele suceder en este tipo entornos, pero no tenemos forma de determinarlo.



- Listado de palabras similares.

| Palabra  | Longitud | Conteo | Porcentaje Ponderado (%) | Palabras Similares              |
|----------|----------|--------|--------------------------|---------------------------------|
| wings    | 5        | 38     | 1,03                     | wings                           |
| today    | 5        | 30     | 0,81                     | today, todays                   |
| thanks   | 6        | 21     | 0,57                     | *thanks, thank, thanks          |
| drinks   | 6        | 17     | 0,46                     | *drink, drink, drinking, drinks |
| energy   | 6        | 14     | 0,38                     | energy                          |
| great    | 5        | 12     | 0,32                     | great                           |
| friends  | 7        | 9      | 0,24                     | *friends, friend, friends       |
| students | 8        | 9      | 0,24                     | student, students               |
| weekend  | 7        | 9      | 0,24                     | *weekend, weekend               |

Tabla 2 – Frecuencia de palabras con conteo igual o superior a 9 del grupo de 423 tweets.

Podemos observar en la Tabla 2, palabras e ideas típicamente asociadas con la marca (energía, alas, bebida, diversión) que describen lo que la marca pretende transmitir y que claramente reproducen sus usuarios en la red, siendo acordes a su imagen de marca.



Por ejemplo, la palabra wings, que sale con mayor frecuencia, está directamente relacionada a la campaña publicitaria de Red Bull "Gives You Wings"<sup>19</sup>, utilizada por la marca. Luego vemos palabras de carácter temporal como today, weekend, morning, tomorrow, friday, monday, etc, indicando que los usuarios hablan de hechos que pasan tanto durante la semana, como en los fines de semana, aunque las referencias al momento presente (today), reciben mayor importancia por parte de los usuarios a la hora de hacer sus publicaciones en Twitter, ya que hay mayor número de menciones

Constatamos también la presencia de palabras que describen el producto como energy y drink, así como palabras relacionadas con lo que proporciona el producto al consumidor, por ejemplo, energised, working, excited, feeling, action, enjoy, needed, inspirational, hydratedetc(ver Anexo 4).

<sup>19</sup> <http://energydrink-us.redbull.com/>



Otra particularidad que podemos ver en el listado, son las palabras relacionadas a los deportes de riesgo patrocinados por la marca, como adventure, vettel, athelte, bikers, force, world, summer, wonder, action, insane, etc.

Por último se puede observar predominio de palabras con significado positivo, como thanks, great, better, friends, amazing, congratulations, lovely, wonder, beautiful, dream, perfect , etc.

A partir del listado de conteo de palabras se generó una nube representativa (Figura 8), con el fin de visualizar mejor la importancia de cada palabra en el contexto general corroborando lo mencionado anteriormente.



Figura 8. Nube de frecuencia de palabras del conjunto de 423 tweets.



Se contempla que las palabras wings y today, tienen mayor tamaño pues aparecen de forma más frecuente. Luego vemos con un tamaño más reducido, aunque también bastante significativo, las palabras thanks, morning, students, inspirational, etc. Lo mismo ocurre a las palabras drinks, energy, enjoy, great, start, commercial etc.

#### 4.1.2.- Etiquetas tipo hashtag (#) más frecuentes

Como mencionado en apartados anteriores, los hashtags son etiquetas que cuando insertadas en forma de (#tema) en el tweet, posibilitan agrupar y filtrar los mensajes. Con base a la muestra de 423 tweets, se generó un listado de frecuencia de palabras, con las 100 etiquetas hashtags según el conteo encontrado (véase tabla 3 y para listado completo consultar Anexo 5).



| Palabra                | Longitud | Conteo | Porcentaje Ponderado (%) |
|------------------------|----------|--------|--------------------------|
| <b>#energy</b>         | 7        | 5      | 0,15                     |
| <b>#spanishgp</b>      | 10       | 5      | 0,15                     |
| <b>#chasingthebull</b> | 15       | 4      | 0,12                     |
| <b>#needtofocus</b>    | 12       | 4      | 0,12                     |
| <b>#addiction</b>      | 10       | 3      | 0,09                     |
| <b>#chasethespirit</b> | 15       | 3      | 0,09                     |
| <b>#cliffdiving</b>    | 12       | 3      | 0,09                     |
| <b>#gopro</b>          | 6        | 3      | 0,09                     |
| <b>#qatar</b>          | 6        | 3      | 0,09                     |
| <b>#rb10</b>           | 5        | 3      | 0,09                     |
| <b>#rbzero</b>         | 7        | 3      | 0,09                     |
| <b>#redbullracing</b>  | 14       | 3      | 0,09                     |
| <b>#sugarfree</b>      | 10       | 3      | 0,09                     |
| <b>#totalzero</b>      | 10       | 3      | 0,09                     |
| <b>#vettel</b>         | 7        | 3      | 0,09                     |

Tabla 3. Etiquetas hashtags más frecuentes de la muestra de 423 tweets.



En la tabla 3 los hashtags **#energy** y **#spanishgp** ocupan los primeros puestos con una frecuencia 5, seguidos por **#chasingthebull** y **#needtofocus** con frecuencia 4, seguidos por las demás etiquetas como **#addiction**, **#chasethespirit**, **#cliffdiving**, **#gopro**, etc, de frecuencia número 3, 2 y 1. Tal información nos revela los temas de más relevancia relacionados a la marca, según los usuarios de Twitter al emplear etiquetas hashtags en sus publicaciones.

Así mismo, se construyó una nube de frecuencia basada en esos hashtags, lo que nos proporcionó un mapa visual y mayor comprensión de los datos (figura 9).





El siguiente paso fue ordenar a los 10 usuarios identificados por su mayor actividad en la red según su número de seguidores, lo que nos revela cuáles son las personas de mayor influencia tienen en la red.

Conforme comentado en el apartado de metodología, con el fin de conservar el anonimato de los usuarios, se les ha cambiado los nombres reales por nombres ficticios. Les llamaremos Usuario A, Usuario B, Usuario C y así sucesivamente.

| Usuarios con más actividad en la red | Tweets publicados | Número de seguidores | Seguidos por la marca |
|--------------------------------------|-------------------|----------------------|-----------------------|
| Usuario A                            | 27                | 157                  |                       |
| Usuario B                            | 18                | 2.269                |                       |
| Usuario C                            | 10                | 868                  |                       |
| Usuario D                            | 10                | 321                  | X                     |
| Usuario E                            | 8                 | 2.136                | X                     |
| Usuario F                            | 7                 | 345                  | X                     |
| Usuario G                            | 7                 | 645                  |                       |
| Usuario H                            | 7                 | 122                  |                       |
| Usuario I                            | 6                 | 1.400                | X                     |
| Usuario J                            | 6                 | 216                  |                       |

Tabla 4. Usuarios con más actividad en la red.



Como puede observarse en las tablas 4 y 5, en relación al número de seguidores, vemos que el Usuario A, a pesar de ser el de mayor actividad, es uno de los que tienen menor número de seguidores. En cambio, el Usuario B, aunque tiene 9 publicaciones menos que el Usuario A, por otro lado es el usuario con mayor número de seguidores de la tabla, seguido por los Usuarios E, I, C, G, F, D, J, A y H. Por lo que podemos concluir que el usuario B es una persona muy influyente.



| Usuarios con más actividad en la red | Tweets publicados | Número de seguidores | Seguidos por la marca |
|--------------------------------------|-------------------|----------------------|-----------------------|
| Usuario B                            | 18                | 2.269                |                       |
| Usuario E                            | 8                 | 2.136                | X                     |
| Usuario I                            | 6                 | 1400                 | X                     |
| Usuario C                            | 10                | 868                  |                       |
| Usuario G                            | 7                 | 645                  |                       |
| Usuario F                            | 7                 | 345                  | X                     |
| Usuario D                            | 10                | 321                  | X                     |
| Usuario J                            | 6                 | 216                  |                       |
| Usuario A                            | 27                | 157                  |                       |
| Usuario H                            | 7                 | 122                  |                       |

Tabla 5. Usuarios con más actividad en la red, ordenados por número de seguidores.

En relación a los usuarios seguidos por la marca, vemos que no hay un seguimiento por parte de la misma a 6 de los 10 usuarios con más actividad en la red, donde 3 de esos, ocupan los 5 primeros puestos del ranking.

#### 4.1.4.- Dispositivos y aplicaciones de publicación

Para interactuar en las redes sociales, los usuarios utilizan distintas maneras de conexión: dispositivos móviles, tabletas, ordenadores de mesa y portátiles dotados de distintos sistemas operativos, aplicaciones, velocidades de conexión y tamaños de pantalla. Identificar y entender esos dispositivos, posibilita a las empresas adaptarse a ellos para que la información llegue a los usuarios de manera efectiva, además de ser una herramienta de alto valor para conocer el perfil de sus usuarios en la red.

Con base a la muestra de 423 tweets, se generó un listado con los dispositivos, sistemas operativos y aplicaciones más utilizadas por los usuarios a la hora de publicar en Twitter (Tabla 6).

| Dispositivo / Aplicación | Longitud | Conteo | Porcentaje Ponderado (%) |
|--------------------------|----------|--------|--------------------------|
| iphone                   | 6        | 386    | 9,32                     |
| instagram                | 9        | 230    | 5,56                     |
| android                  | 7        | 108    | 2,61                     |
| facebook                 | 8        | 24     | 0,58                     |
| hootsuite                | 9        | 20     | 0,48                     |
| percolate                | 9        | 20     | 0,48                     |
| tweetdeck                | 9        | 14     | 0,34                     |
| crowdbooster             | 12       | 8      | 0,19                     |
| twitterfeed              | 11       | 8      | 0,19                     |
| Apple - OSX              | 5        | 4      | 0,10                     |
| foursquare               | 10       | 4      | 0,10                     |
| windows                  | 7        | 4      | 0,10                     |
| blackberry               | 10       | 3      | 0,07                     |
| smalltalkmessaging       | 18       | 3      | 0,07                     |
| socialflow               | 10       | 2      | 0,05                     |
| bufferapp                | 9        | 1      | 0,02                     |
| itunes                   | 6        | 1      | 0,02                     |
| publicize                | 9        | 1      | 0,02                     |
| sproutsocial             | 12       | 1      | 0,02                     |
| tablets                  | 7        | 1      | 0,02                     |
| wordpress                | 9        | 1      | 0,02                     |

Tabla 6. Dispositivos y aplicaciones de publicación.



Como podemos observar en la tabla 6, el mayor número de publicaciones proviene con diferencia desde dispositivos móviles, modelos iPhone de la marca Apple<sup>20</sup>. Ocupando el segundo puesto, aparece la aplicación de fotos Instagram<sup>21</sup> también diseñada para usarse mayoritariamente en dispositivos móviles.

En tercer puesto seguimos en el campo de los dispositivos móviles, pero ahora dotados del sistema operativo Android<sup>22</sup>. Vemos también otras redes sociales interactuando con Twitter, como Facebook<sup>23</sup> y foursquare<sup>24</sup>.

<sup>20</sup> <http://www.apple.com/about/>

<sup>21</sup> <http://instagram.com/about/faq/>

<sup>22</sup> <http://source.android.com/>



También encontramos un considerable número de API's de gestión de cuentas en Twitter, como Hootsuite<sup>25</sup>, Tweetdeck<sup>26</sup>, twitterfeed<sup>27</sup>, etc. El listado también nos revela publicaciones realizadas por sistemas operativos como OSX y Windows además de otros dispositivos móviles como BlackBerry y tabletas.

A partir de la información de la Tabla 6, se creó una nube de frecuencia de palabras a fin de visualizar mejor los resultados, donde los dispositivos y aplicaciones más usados como iphone, instagram, android, cobran mayor tamaño en la nube (Figura 10).



Figura 10. Nube de palabras de aplicaciones fuente.



#### 4.1.5. - Mapa de geolocalización

Twitter también proporciona al usuario que posea un dispositivo geolocalizable, la posibilidad de enviar esos datos en formato de latitud y longitud, como información añadida al tweet. Eso posibilita saber la localización geográfica del usuario en el momento que realizó la publicación del mensaje.

---

<sup>23</sup> <https://www.facebook.com/facebook?brandloc=DISABLE>

<sup>24</sup> <https://es.foursquare.com/about>

<sup>25</sup> <https://hootsuite.com/about>

<sup>26</sup> <https://about.twitter.com/es/products/tweetdeck>

<sup>27</sup> <http://help.twitterfeed.com/knowledgebase/articles/88061-what-is-twitterfeed>



Con base a la muestra de 423 tweets, se identificaron 90 tweets con información de latitud y longitud, lo que hizo posible generar un mapa de geolocalización de los mismos (Figura 11).



Figura 11. Mapa de geolocalización de la muestra de 423 tweets.

En el mapa arriba, vemos que la mayor concentración de tweets proviene de zonas consideradas de alto desarrollo económico como América del Norte y Europa, seguidos por Centro América, América del Sur, África del Sur, Australia, Nova Zelanda y Asia con Indonesia, Japón, Filipinas, China y Corea del Sur.

Otro punto interesante a destacar es la gran concentración de tweets generados desde las zonas de costa marítima. El punto número 6 además, muestra que el usuario estaba en alta mar cuando hizo su publicación en Twitter.

#### 4.1.6. – Ranking retweets

El número de retweets (RT) de un tweet, es un indicador de “éxito” del mismo en la red, pues cuando retweeteas un tweet, éste aparece en tu muro y todos tus contactos podrán verlo, ampliando el alcance del mensaje.



También está la opción de marcar ese tweet como favorito, lo que da al usuario la posibilidad de acceder a ese tweet en otro momento en su lista de favoritos. De modo que si un tweet tiene un alto número de RT's y además ha sido marcado como favorito una gran cantidad de veces, todo indica que este tweet tiene un alto grado de importancia en la red. En la tabla 7, elaboramos con base al conjunto de datos de 423 tweets, un ranking de los 10 tweets con mayor número de RT's con el fin de identificar los tweets de mayor éxito dentro de la muestra. La tabla está compuesta por el tweet, número de retweets y número de veces que ha sido marcado como favorito.

| Tweet  | RT | Favorito |
|--|----|----------|
| Just had a kind of supercopter day :-) What a ride !! *LotOfFun *GivesYouWings<br><a href="http://t.co/BSuK5TwJIB">http://t.co/BSuK5TwJIB</a>  | 42 | 54       |
| Actual Help Desk Time! Come find me on the second floor of Hale for some Red Bull<br>*givesyouwings  | 9  | 3        |
| finals got you down? come find me on the second floor of Hale for some free ice cold<br>*RedBull *givesyouwings  | 7  | 3        |
| Last Actual Help Desk of the year! Come find me near the printers in Hale for your<br>*FinalsWings *RedBull *SugarFree *GivesYouWings  | 7  | 0        |
| Dive into the new week! *GivesYouWings <a href="http://t.co/n6wZk3yIRw">http://t.co/n6wZk3yIRw</a>   | 5  | 4        |
| Today @redbullire blew google Ireland's minds with the oculus *cliffdiving *givesyouwings<br><a href="http://t.co/3I00zqnd8s">http://t.co/3I00zqnd8s</a>   | 5  | 4        |
| â€œ@redbull: Redefine your definition of self. @denitiaandsene *givesyouwings<br><a href="http://t.co/b22MJPHBZz">http://t.co/b22MJPHBZz</a> <a href="http://t.co/PwZITU7I8Iâ€;@DetroitChe">http://t.co/PwZITU7I8Iâ€;@DetroitChe</a> | 5  | 1        |
| Red Bull gives you balls (from a girl to a boy) @vikkram_ *redbullsg *givesyouballs<br>*givesyouwings <a href="http://t.co/EjG8aYMwze">http://t.co/EjG8aYMwze</a>  | 4  | 7        |
| Congratulations to @AndrewJordan77 @olsbergsmse for the @RallycrossRX podium *redbull<br>*givesyouwings *achievement <a href="http://t.co/dHsRbeclCZ">http://t.co/dHsRbeclCZ</a>   | 3  | 16       |

Tabla 7. Ranking de tweets, según RT's.

Se observa que el primer tweet "Just had a kind of supercopter day :-) What a ride !!\*LotOfFun \*GivesYouWings <http://t.co/BSuK5TwJIB>" se destaca con diferencia de los demás en cuanto a la cantidad de RT's. La foto (figura 12) fue publicada en Twitter y nos deja a entender, que el usuario voló en un helicóptero patrocinado por Red Bull.





Figura 12. Foto publicada por @Usuario en Twitter.

Fuente: <http://t.co/BSuK5TjJlB>

#### 4.2.– Análisis detallado de la muestra aleatoria de 100 tweets

cim

Como comentado anteriormente, uno de los campos de estudio de la Minería de Opiniones, es el de la identificación de la polaridad de la información, clasificándola como positiva, negativa y neutral.

En este apartado, presentaremos los resultados obtenidos a través de la evaluación subjetiva de la polaridad realizada a la muestra aleatoria de 100 tweets, que fueron divididos en dos grupos de 50 tweets, llamados Red Bull1 y Red Bull2 y evaluados por un total de 109 individuos. Como resultado se generaron 4 subgrupos de datos:

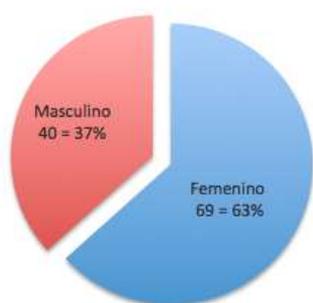
- Tweets de valencia positiva.
- Tweets de valencia neutro-positiva.
- Tweets de valencia neutro-negativa.
- Tweets de valencia negativa.

##### 4.2.1.– Análisis descriptivo de la encuesta

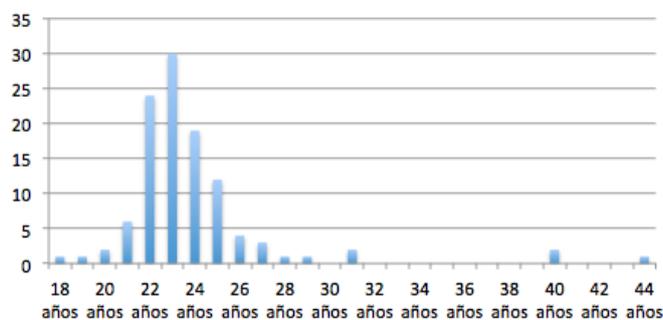


Las encuestas fueron aplicadas a 109 estudiantes de la Universidad de Granda que conocen la marca. Concretamente, la muestra está compuesta por hombres (37%) y mujeres (63%) de edades comprendidas entre 18 y 44 años y en su mayoría, con educación superior, como puede observarse en los siguientes gráficos. Además, un 46% de la muestra consume Red Bull.

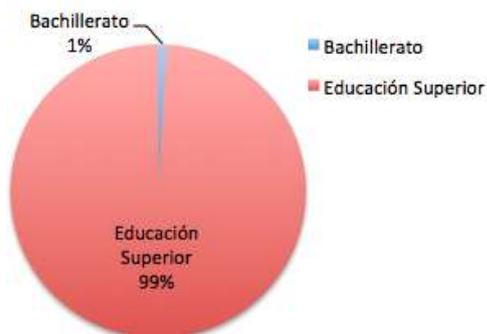
**Sexo**



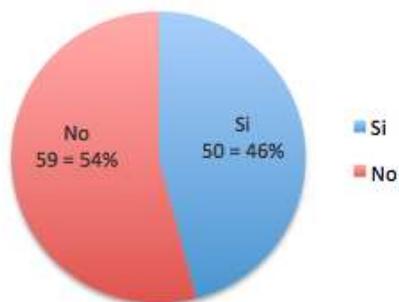
**Edad**



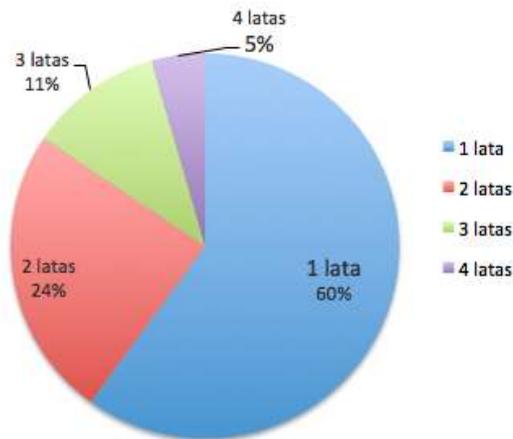
**Nivel de Estudios**



**Consume Actualmente Red Bull?**



### Consumo medio a la semana, si consume Red Bull



#### 4.2.2 – Tweets de valencia positiva

En su conjunto, de los 100 tweets evaluados, 68 fueron clasificados como positivos (Anexo 6). En dicho grupo de tweets, se encuentran mensajes que expresan una actitud muy positiva hacia la marca. Mensajes que hacen referencia a agradecimientos por lo que le proporciona el producto, comentarios sobre la publicidad realizada por la marca, sentimientos del cliente al consumir el producto, la necesidad de consumir el producto, la preferencia por una línea específica del producto “Silver”, etc.



Algunos ejemplos de estos tweets son los siguientes:

- **Absolutely love red bull <3 #GivesYouWings.**
- **@redbull @redbullOHIO My office thanks you! #wiiings #givesyouwings.**
- **Coming for a party cave! Get excited for some wings #givesyouwings #redbull @redbulldfw @redbullbigt.**
- **Red Bull has some of the best commercials. #GivesYouWings.**



- Thanks to @redbull for keeping me up to study for finals #finalweek #givesyouwings.
- Thanks you red bull for getting me truth work #givesyouwings #RedBull.
- Mmm I love the smell of a red bull when I first crack it open on a Friday afternoon. #LikeClockwork #redbull #givesyouwings #fridaynight.
- Takin' ma wings for a walk! #walkthewalk #swagger #performamce #givesyouwings #girdegvinger.
- Enjoying the sun at my balcony after todays workout!! @redbull #givesyouwings #TiMetteLeAli.
- Need this today. #redbull #givesyouwings.
- Every Red Bull commercial never disappoints. #sogood #GivesyouWings.
- My body is exhausted but my mind is racing. Thankyou red bull.... #givesyouwings #idontwantwingsrightnow
- Nothing better than a crisp Silver Edition. #givesyouwings @ AngelFire Resort.
- The sky's the limit. #givesyouwings #redbull.
- It's a redbull type of day! #givesyouwings.

cim

Sobre este conjunto de 68 tweets positivos, se realizaron los filtrados de frecuencia de palabras, revelando que las palabras today, thanks, wings, drink, start, better, ocupan los primeros puestos de mayor importancia corroborando la connotación positiva asociada al grupo (véase tabla 8 y para listado completo consultar Anexo 7).



| Palabra     | Longitud | Conteo | Porcentaje Ponderado (%) | Palabras Similares      |
|-------------|----------|--------|--------------------------|-------------------------|
| today       | 5        | 7      | 2,33                     | today, todays           |
| thanks      | 6        | 6      | 2,00                     | #thanks, thank, thanks  |
| wings       | 5        | 6      | 2,00                     | wings                   |
| drink       | 5        | 5      | 1,67                     | drink, drinking, drinks |
| start       | 5        | 4      | 1,33                     | start, starting         |
| better      | 6        | 3      | 1,00                     | better                  |
| commercials | 11       | 3      | 1,00                     | commercial, commercials |
| friends     | 7        | 3      | 1,00                     | friend, friends         |
| shine       | 5        | 3      | 1,00                     | shine, shining          |
| tomorrow    | 8        | 3      | 1,00                     | tomorrow                |
| weekend     | 7        | 3      | 1,00                     | weekend                 |

Tabla 8. Frecuencia de palabras de los 68 tweets positivos.

Así mismo, se generó una nube de frecuencia de palabras (Figura 13) a partir de estos datos donde se visualiza por un lado, palabras como thanks, today, wings, drinks y friends, que tienen mayor tamaño por su alta frecuencia. Por otro lado, aunque varían en frecuencia, otras palabras que también resaltan son:



- Better, tomorrow y commercials, etc.
- Start, weekend, enjoy, great, friday, feeling, etc.
- Shine, inspirational, first, getting, etc.







Figura 14. Nube de frecuencia de hashtags de los 68 tweets positivos.

#### 4.2.3 – Tweets de valencia neutro-positiva

El grupo de los tweets de valoración neutro-positiva está compuesto por 24 mensajes (véase anexo 9). Se observa una connotación más neutra en relación a la marca, aunque sigue siendo positiva, relacionada sobre todo al carácter energético de la bebida y a su utilidad en la vida cotidiana de los usuarios. También destaca la referencia a una línea específica del producto, sin azúcar, “Red Bull Total Zero”



Algunos ejemplos son:

- can't go through school without it \*redbull \*givesyouwings
- Race ready with \*RedBullTotalZero at Brentwood ski and snowboard centre \*givesyouwings.
- Just remembered I can go back to drinking redbulls \*junky \*givesyouwings
- Good Luck in your exams \*UEL Students!! Come to the campus shop and grab an energy drink!  
\*GivesYouWings \*ExamProblems.
- Football tonight so some much needed artificial energy is in order \*redbull  
\*givesyouwings



En el análisis de la frecuencia de palabras de este grupo, se nota un número reducido de palabras en relación a los listados de frecuencia de la muestra anterior. Ese fenómeno viene explicado por el pequeño tamaño de la muestra, 24 tweets.

En el grupo de tweets neutro-positivos (véase tabla 10 y para listado completo Anexo 10), la palabra wings asume uno de los últimos puestos del listado de frecuencia y la palabra students, aparece en primer puesto, seguida por las palabras, drink, energised, energy, exams y today. Se puede denotar una relación entre el tipo de usuario del producto y el uso que tal usuario hace del mismo. Esto es, se puede concluir que los estudiantes utilizan la bebida energética como apoyo durante el periodo de exámenes, lo que explica la connotación neutro-positiva del grupo.

| Palabra   | Longitud | Conteo | Porcentaje Ponderado (%) | Palabras Similares    |
|-----------|----------|--------|--------------------------|-----------------------|
| students  | 8        | 3      | 2,29                     | students              |
| drink     | 5        | 2      | 1,53                     | drink, drinking       |
| energised | 9        | 2      | 1,53                     | energised, energising |
| energy    | 6        | 2      | 1,53                     | energy                |
| exams     | 5        | 2      | 1,53                     | exams                 |
| today     | 5        | 2      | 1,53                     | today                 |

Tabla 10. Frecuencia de palabras de los 24 tweets neutro-positivos.

A partir de estos datos, se generó una nube representativa de frecuencia de palabras (Figura 15), con el fin de visualizar mejor la importancia de cada palabra en el contexto general.



Su connotación neutro-negativa viene explicada por las palabras, break, confort, droppin, never y nigga<sup>28</sup>, expresión utilizada para referirse a la raza de color negra en la lengua inglesa.

Igualmente, debido al número reducido de tweets, el listado de frecuencia de palabras también se reduce (véase tabla 10).

| Palabra | Longitud | Conteo | Porcentaje Ponderado (%) | Palabras Similares |
|---------|----------|--------|--------------------------|--------------------|
| break   | 5        | 1      | 8,33                     | break              |
| comfort | 7        | 1      | 8,33                     | comfort            |
| droppin | 7        | 1      | 8,33                     | droppin            |
| never   | 5        | 1      | 8,33                     | never              |
| nigga   | 5        | 1      | 8,33                     | nigga              |

Tabla 11. Frecuencia de palabras de los tweets neutro-negativos.

Teniendo en cuenta el diminuto tamaño de la muestra y el bajo valor de conteo de frecuencia de palabras, la nube de palabras pierde su función. Por ello, en este caso, no se ha generado la nube de palabras.

#### 4.2.5- Tweets de valencia negativa

El cuarto y último grupo está formado por 5 tweets con polaridad negativa según la evaluación subjetiva. Este grupo de publicaciones, por un lado hace referencia a los efectos negativos del exceso de consumo del producto. Por otro lado se puede detectar el relato de un usuario que echa en falta una línea del producto que ya no está en venta (Anexo 12). Por ejemplo:

- So much caffeine, I feel like a squirrel jacked on red-bull #givesyouwings #caffeinejitters.
- So that redbull for dinner was definitely a bad idea.. #WideAwake #GivesYouWings #TeamNoSleep.

<sup>28</sup> <http://en.wikipedia.org/wiki/Nigga>



- J-spec 6 oz. Red Bull... I miss the 12 oz. Total Zero :P \*redbull \*givesyouwings  
\*lifeofstickydijoe.

Según el listado de frecuencia, se concluye que las palabras que caracterizan este grupo son de connotación negativa y temporal (tabla 12). Por un lado, tenemos caffeine y jacket que podemos calificarlas como negativas y por otro lado, las temporales, today y tomorrow.

| Palabra  | Longitud | Conteo | Porcentaje Ponderado (%) | Palabras Similares |
|----------|----------|--------|--------------------------|--------------------|
| Caffeine | 8        | 1      | 3,03                     | Caffeine           |
| Jacked   | 6        | 1      | 3,03                     | Jacked             |
| Today    | 5        | 1      | 3,03                     | Today              |
| Tomorrow | 8        | 1      | 3,03                     | Tomorrow           |
| Total    | 5        | 1      | 3,03                     | Total              |

Tabla 12 – Frecuencia de palabras de los tweets negativos.

Del mismo modo que en los casos anteriores, debido al número reducido de tweets, el listado de frecuencia de palabras se reduce, y la nube de frecuencia de palabras pierde su función, por lo cual para este grupo no se ha generado la nube de palabras.



#### 4.2.6 – Análisis de los emoticonos

En relación a los emoticonos encontrados en la muestra de 100 tweets, se ha hecho un análisis buscando identificar los diferentes tipos y su relación con la evaluación subjetiva realizada.

En el grupo de tweets de valencia positiva, se identifican emoticonos que representan un rostro sonriente y corazones, lo que indica congruencia con la evaluación positiva recibida (Tabla 12).



| Tweet  | Emoticono |
|--|-----------|
| @user: Just had a kind of supercopter day :-) What a ride !! *LotOfFun<br>*GivesYouWings                         | 😊         |
| Absolutely love red bull <3 *GivesYouWings   | 💙         |
| Today's fun :) *skate *funday *maythefourthbewithyou *dc *redbull *unit23<br>*hashtag ;P                         | 😊 🤪       |
| My friends are so fly! <3 *teamredbull @redbull *givesyouwings @jeff_provenzano<br>@RedBullAirForce @user        | 💙         |
| Effort sometimes is rewarded..<3 <3 *RedBull *shirt from *Team *thanks *sosweet<br>*shirt *loveit *givesyouwings | 💙 💙       |
| Giving wings to the bikers *ChaseTheSpirit *givesyouwings What a click !! :-)@redbull                            | 😊         |

Tabla 12. Lista emoticonos de los 68 tweets positivos.

En el grupo de tweets de valencia neutro-positiva, se identifica apenas un emoticono. Éste indica un rostro sonriente, lo que corrobora la evaluación neutro-positiva de la muestra (Tabla 13).

cim

| Tweet  | Emoticono |
|--|-----------|
| @user: Just had a kind of supercopter day :-) What a ride !! *LotOfFun<br>*GivesYouWings | 😊         |

Tabla 13. Lista emoticonos de los 24 tweets neutro-positivos.

Para la muestra de tweets de valencia neutro-negativa, no se encontraron emoticonos en los tweets.

Referente a la muestra de tweets de valencia negativa, se identifica apenas un emoticono. Éste indica un rostro con la lengua hacia fuera, que en este caso tiene connotación negativa, pues el mensaje indica un reclamo por parte del cliente de un producto con más cantidad (12oz) y sin azúcar



(Total Zero). Lo que podemos afirmar que también es acorde con la evaluación subjetiva realizada (Tabla 14).

| Tweet  | Emoticono   |
|--|---|
| J-spec 6 oz. Red Bull... I miss the 12 oz. Total Zero :P #redbull #givesyouwings #lifeofstickydiljoe |  |

Tabla 14. Lista emoticonos 5 tweets negativos.

#### 4.3. – Análisis automático-objetivo de los datos

Se realizó un análisis automático-objetivo basado en la muestra de 100 tweets, con el fin de comparar las coincidencias entre la evaluación subjetiva realizada anteriormente y la evaluación automática-objetiva realizada a través de la herramienta R (véase tabla 15 y para listado completo en Anexo 13). Para ello, se utilizaron 3 medidas de valoración, (positiva, neutra, negativa) de modo que los tweets evaluados como neutro-positivos y neutro-negativos fueran considerados como neutros (zona neutral) al momento de realizar la comparación entre los dos métodos.

| Tweet  | Categoría SAM   | Polaridad con SAM | Categoría R | Polaridad con R | Coincidencia |
|--|-----------------|-------------------|-------------|-----------------|--------------|
| @redbull Just what you need on a hot day at the beach<br>#Givesyouwings @RedBullUAE <a href="http://t.co/z2b8tOMP7d">http://t.co/z2b8tOMP7d</a>                                | Positivo        | 3,69              | Positivo    | 1               | Sí           |
| Good Luck in your exams *UEL Students!! Come to the campus shop and grab an energy drink! *GivesYouWings<br>*ExamProblems  | Neutro-Positivo | 4,20              | Positivo    | 2               | No           |
| @redbullNYC: Rise and shine like there's no tomorrow.<br>#givesyouwings  | Positivo        | 2,85              | Positivo    | 2               | Sí           |
| "@redbull: Break out of your comfort zone.<br>#givesyouwings <a href="http://t.co/1k80pqV7n">http://t.co/1k80pqV7n</a> " boutta kick the shit out of that nigga in the back.   | Neutro-Negativo | 5,76              | Negativo    | -1              | No           |
| Thank you Red Bull ladies for the wonderful beverage at TOPP Drift today! *givesyouwings #redbull<br><a href="http://t.co/gLRQBtVGDB">http://t.co/gLRQBtVGDB</a>               | Neutro-Positivo | 4,35              | Positivo    | 2               | No           |
| @redbullIRE had a great time giving wings to kitesurfers at the Clontarf *BattleForTheBay today! *givesyouwings<br><a href="http://t.co/SvT72Fqocg">http://t.co/SvT72Fqocg</a> | Positivo        | 3,47              | Positivo    | 1               | Sí           |
| New plan. Drink nothing but @redbull all weekend.<br>#Givesyouwings #soberweekend  | Postivo         | 3,71              | Neutro      | 0               | No           |
| Every Red Bull commercial never disappoints. *sogood<br>#GivesyouWings   | Postivo         | 3,35              | Negativo    | -1              | No           |
| Football tonight so some much needed artificial energy is  | Neutro-Postivo  | 4,29              | Neutro      | 0               | Sí           |



|  |          |      |          |   |    |
|--|----------|------|----------|---|----|
| in order #redbull #givesyouwings...<br><a href="http://t.co/UL28T2tPEY">http://t.co/UL28T2tPEY</a> |          |      |          |   |    |
| @acoona_matataa I prefer redbull! #GivesYouWings   | Positivo | 3,60 | Positivo | 1 | Sí |

Tabla 15. Comparativa análisis subjetivo y objetivo 100 tweets.

Al analizar los resultados de la comparación entre los distintos métodos, se concluye que las valoraciones coinciden apenas un 49% de las veces, lo que indica un muy bajo nivel de coincidencia entre las valoraciones subjetivas realizadas por personas y las objetivas automáticas realizadas por ordenadores. A continuación citamos algunos ejemplos en los que la evaluación fue muy coincidente entre ambos métodos (Tabla 16).

| Tweet   | Categoría SAM   | Polaridad con SAM | Categoría R | Polaridad con R | Coincidencia |
|---|-----------------|-------------------|-------------|-----------------|--------------|
| Have all a good run tomorrow. Enjoy it! #wingsforlife<br>#worldrun #givesyouwings #spielberg<br><a href="http://t.co/DcEF59i11E">http://t.co/DcEF59i11E</a> | Positivo        | 2,80              | Positivo    | 2               | Sí           |
| So that redbull for dinner was definitely a bad idea..<br>#WideAwake #GivesYouWings #TeamNoSleep  | Negativo        | 6,44              | Negativo    | -1              | Sí           |
| I wish @redbull came in a glass bottle. #givesyouwings  | Neutro-Negativo | 5,39              | Neutro      | 0               | Sí           |

Tabla 16. Tweets con mayor grado de coincidencia de polaridad.

Como podemos observar en la tabla 16, hay que resaltar el caso del tweet que ocupa el tercer puesto "I wish @redbullcame in a glassbottle. #givesyouwings", clasificado en la categoría neutral. Este tweet representa un deseo o reclamo por parte del cliente de un producto con envase no utilizado por Red Bull actualmente, un envase de cristal. La evaluación humana subjetiva lo interpreta como neutro-negativo, pues entiende que el cliente no está descontento con el envase actual del producto, no en tanto preferiría mejor un envase de cristal. Lo que justifica tal evaluación subjetiva.

A continuación citamos algunos ejemplos en los que la evaluación fue muy divergente entre ambos métodos (Tabla 17).

| Tweet  | Categoría SAM | Polaridad con SAM | Categoría R | Polaridad con R | Coincidencia |
|--|---------------|-------------------|-------------|-----------------|--------------|
| Mmm I love the smell of a red bull when I first crack it<br>open on a Friday afternoon. #LikeClockwork #redbull<br>#givesyouwings #fridaynight | Positivo      | 2,20              | Negativo    | -1              | No           |
| Today just hasn't been a good day for me. Need to make<br>the best of tomorrow. #givesyouwings   | Negativo      | 6,08              | Positivo    | 2               | No           |

Tabla 17 - Tweets con mayor grado de divergencia de polaridad.



Se observa un dato muy interesante en la tabla 17, el tweet "Today just hasn't been a good day for me. Need to make the best of tomorrow. #givesyouwings", cuando evaluado por el algoritmo automático, recibe la valoración positiva, pues lleva en el mensaje las palabras "good" y "best", sin embargo en la valoración subjetiva (SAM), recibe valoración negativa, pues se tiene en cuenta el sentido general de la frase.



## 5. Conclusiones

---

Caracterizada por la creación colaborativa de contenido (Lévy, 2004) y permitiendo la interacción social y diseminación de información de manera rápida y sencilla (Kaplan & Haenlein, 2010), las redes sociales en la web 2.0 presentan nuevas perspectivas a los clientes y empresas, pues posibilitan a los consumidores expresar sus opiniones en relación a productos y servicios en la red (Hunt, 2010). Las redes sociales actúan por lo tanto como un barómetro, midiendo cómo una organización es vista y sentida por sus clientes.

Teniendo la interacción como uno de sus elementos estructurales principales, las redes sociales representan por tanto un importante canal de comunicación para la empresa con sus clientes, en un entorno donde el consumidor gana voz y el diálogo pasa a ser directo y constante. Tal entorno nutre a las empresas que mantienen un enfoque de marketing relacional, ya que aporta información extremadamente relevante que posibilita entender mejor a los clientes, lo que necesitan y esperan de la empresa (Kotler & Armstrong, 2008).

En el intento de identificar y estudiar la orientación de las opiniones expresadas por los usuarios en la red, la Minería de Opiniones y Análisis de Sentimientos surgen como un conjunto de técnicas que permiten recabar, procesar y comprender grandes cantidades de información (Chen & Zimbra, 2010), en busca de insights y relatos que sean importantes para ambas partes, posibilitando avances en la relación con el cliente además de mejoras de los servicios o productos ofrecidos.

En este contexto, el presente trabajo tiene como objetivo aplicar las técnicas de Minería de Opiniones y Análisis de Sentimientos a un conjunto de publicaciones relacionadas con la marca Red Bull, creadas por los usuarios en la red social Twitter bajo la etiqueta (*#givesyouwings*), con el fin de reunir e interpretar datos que puedan ser utilizados por las empresas en la aplicación de estrategias de marketing relacional.

De forma general, los resultados obtenidos apuntan que los usuarios en su mayoría hablan positivamente de Red Bull en la red, cuando palabras como *thanks, love, energy, great, enjoy, better* y *awesome*, aparecen con una alta frecuencia en sus publicaciones en Twitter.

En la misma línea se identificaron también un elevado número de mensajes de tono positivo y de gratitud, publicados por estudiantes que hacen uso del producto en períodos de exámenes.

Otro dato significativo observado hace referencia a sugerencias y propuestas realizadas por los clientes en la web. Así, destacan mensajes que contienen sugerencias de mejora del producto, como por ejemplo, cambios en el envase, así como peticiones reivindicando líneas o categorías del producto que ya no se comercializan.

Los análisis realizados a partir de los tweets extraídos, permitieron identificar también las cuentas de usuarios más activos en la red. Además, gracias al modelo de seguidores y seguidos que proporciona Twitter, se pudo elaborar un ranking de esos usuarios, lo que determina el grado de influencia del mismo en la red. A las empresas con enfoque relacional, tal calidad de información se hace imprescindible para las prácticas de marketing directo.

Referente a la geolocalización extraída de los datos, se aprecia una baja adhesión en la utilización del servicio por parte de los usuarios (16,91%), por lo cual se recomienda a las empresas estimular la utilización del servicio de geolocalización junto a sus clientes. Con tal información, las empresas pueden ver dónde se encuentran los usuarios cuando publican sus mensajes y en qué zonas del globo están ubicados sus clientes, lo que posibilita entender aspectos culturales, políticos, económicos y geográficos de los mismos, como por ejemplo, costumbres, conocimientos, grado de desarrollo artístico, científico y industrial, religión, moneda, PIB, tasa de importación y exportación, clima, flora, fauna, hidrografía, etc. Tal conocimiento es de suma importancia a la hora de entender el entorno en que vive el cliente, y facilita a las corporaciones crear y ofrecer productos y servicios acorde a las necesidades reales del cliente, así como un mayor ajuste en base a las características del entorno en el que se van a desarrollar las prácticas de marketing.

Finalmente, como resultado de las dos diferentes técnicas de Análisis de Sentimientos utilizadas en el trabajo, evaluación subjetiva a través de la escala pictográfica SAM y algoritmos automáticos realizados a través de la herramienta R, se concluye que tienen un bajo grado de coincidencia entre sí (49%). Este dato sugiere lo susceptible que es el cálculo automático, pues depende de algunos factores importantes como, lo bueno que es el diccionario que se utiliza, o por ejemplo, si tiene incorporado valencia para los emoticonos. Otro punto a destacar, es que la evaluación automática aplica un análisis restringido centrado en las palabras del texto de manera individual y aislada, sin tener en cuenta la frase completa, lo que se puede traducir en errores a la hora de valorar mensajes que llevan tono de ironía o sarcasmo, corroborando lo que dicen Pang & Lee (2008) en su estudio

sobre el Análisis de Sentimientos a través de los algoritmos automáticos.

Resultados muy parecidos encontró la empresa Campofrío en su estudio de la campaña "Hazte extranjero", analizando los tweets que contenían la palabra "Campofrío" y el hashtag #HazteExtranjero en Twitter. Los resultados de este estudio apuntan por un lado que los algoritmos automáticos agilizan el proceso de análisis, pero que por otro lado, los resultados encontrados por este proceso automático discrepan a los resultados conseguidos de forma manual (Solo Marketing, 2014).

En definitiva, los resultados indican una baja fiabilidad en el análisis a través de los algoritmos automáticos, que viene explicada por el bajo nivel de coincidencia encontrado (49%), cuando se compara el método automático al subjetivo. Según la literatura, esos errores provienen de la dificultad en analizar e interpretar el mensaje como un todo, y no apenas palabra por palabra separadamente. Que una palabra positiva o negativa aparezca en un tweet, no necesariamente indica que el tweet en su conjunto sea positivo o negativo. Esta es una limitación importante que tiene el método de conteo de palabras y que no tiene el método a través de evaluaciones subjetivas, ya que las personas evalúan la polaridad del tweet en su conjunto. Tal deficiencia hace que los mensajes con tono de ironía y sarcasmo, no sean bien clasificados por los algoritmos automáticos

En resumen, los temas abordados y el extenso abanico de técnicas utilizadas en el presente trabajo, adquieren singular relevancia para las empresas en el panorama actual, donde las redes sociales representan además de un canal directo de dialogo con el cliente, un pozo repleto de valiosa información a respecto del comportamiento, perfil geográfico, económico y cultural de los mismos. Basta con que las empresas sepan recabar, filtrar y entender tal "elixir" para que puedan actuar de manera efectiva y acertada, en su misión de diseñar y entregar la mejor propuesta a cada tipo de cliente.

cim



## 6. Limitaciones

---

A continuación se describen algunas de las limitaciones tanto de carácter temporal como metodológico que presenta este trabajo. Concretamente, son las siguientes:

- **Limitaciones en la descarga de los datos.**

Una dificultad importante estuvo asociada al hecho de obtener una descarga más amplia de los datos, pues se tuvo que hacer varios intentos al día para conseguir un número muy reducido de tweets. Aunque según la herramienta R se pueden descargar más de 1500 tweets de una sola, aunque este número es difícil de alcanzar debido a que Twitter limita la descarga y muchas veces repite los datos. De hecho, para conseguir la muestra de datos utilizada en este trabajo, se necesitaron 23 días. Por lo tanto, asociada a esta dificultad en la descarga de datos, una limitación de este trabajo está en el número final de datos conseguidos; de no haber encontrado esta dificultad se hubiese trabajado con conjuntos de datos más grandes y por lo tanto resultados más representativos.

- **Limitaciones referentes a la red social.**

Este trabajo se centra en la información contenida bajo el #gisvesyouwings, en la red social Twitter. No se han tenido en cuenta otras redes sociales que también son de suma importancia en el escenario web (por ejemplo Facebook) que también trabajan con la etiqueta estudiada y pueden ser una importante fuente de datos para el estudio.

- **Limitación referente al idioma.**

El trabajo se ha centrado en estudiar los tweets generados en el idioma inglés. No obstante, Red Bull es una marca que está presente en diversos países y se comunica con sus seguidores en diversos idiomas. Por ejemplo, en los países de habla hispánica, la marca trabaja bajo el hashtag (#tedaalas), en los países de habla portuguesa(#tedaasas), en países de habla

alemana (\*verleihtFlügel), por lo que se asume la necesidad de ampliar la investigación a otros idiomas, a fin de tener una idea más global de qué hablan y piensan los clientes de la marca en las redes sociales.

- **Limitación relacionada al estudio de los emoticonos.**

Este trabajo hace un breve análisis de los emoticonos encontrados en la muestra aleatoria de 100 tweets, creada para la evaluación subjetiva, pero no se realiza un análisis más específico sobre el tema. No obstante, dada la importancia adquirida por los emoticonos en los mensajes publicados en la web, se hace patente la necesidad de realizar estudios más en profundidad de los mismos.

Otro punto a destacar, es la ausencia de un análisis automático basado en los emoticonos encontrados en los mensajes. Un análisis de tal naturaleza serviría como un instrumento más, para medir y comparar las coincidencias entre los métodos automáticos y subjetivos

- **Limitaciones sobre el manejo de los datos.**

Los datos han sido manejados en gran medida manualmente, en especial con las herramientas Excel y Mapquest. Teniendo en cuenta el gran volumen de información utilizada en esta investigación, carecer de herramientas que automaticen ese manejo, supone una limitación importante en términos de atención y tiempo por parte del investigador.

- **Limitaciones en relación a la muestra aleatoria de 100 tweets.**

Debido al factor tiempo, se optó por evaluar subjetivamente apenas 100 tweets, muestra aceptable estadísticamente, aunque una muestra mayor proporcionaría un estudio más completo y resultados aún más robustos.

- **Limitaciones de la encuesta.**

Otra limitación importante está relacionada a la encuesta que fue aplicada a 109 estudiantes de la Universidad de Granada, que evaluaron un grupo de 50 tweets cada uno. Por un lado se presenta la limitación referente al número de individuos que aunque supera el

mínimo requerido estadísticamente, un número mayor de individuos ayudaría a reforzar los resultados obtenidos.

Por otro lado, la muestra se compone solamente por estudiantes de la Universidad de Granada, lo que limita el estudio. Una muestra más diversa aumentaría la relevancia de los datos.

- **Limitación referente a la geolocalización.**

Debido al número tan reducido de usuarios de la red que tienen la opción de geolocalización activa en su dispositivo, la muestra resultó relativamente pequeña para un análisis más concluyente.

.  
.



## 7. Futuras Líneas de Investigación

---

No obstante, muchas de estas limitaciones constituyen líneas de actuación futuras. A continuación presentamos algunas recomendaciones como futuras líneas de investigación, a título de complementar más el estudio realizado. Concretamente, se plantean las siguientes ideas:

- **Ampliar la muestra de datos extraída del la web de dos maneras:**
  - **Extendiendo la búsqueda de información a otras redes sociales que posibiliten la utilización de etiquetas del tipo hashtags.**
  - **Abarcando una mayor diversidad de idiomas al estudio.**
- **Se sugiere también un análisis pormenorizado de las características de la muestra con el fin de identificar detalladamente el perfil de esos individuos, según cada grupo de datos conseguidos a través del análisis subjetivo.**
- **Realizar un estudio en profundidad enfocado a los retweets y tweets marcados como favoritos, para entender la relevancia en términos de contenido y el “alcance” de un tweet.**
- **Estudiar a los tweets y retweets publicados por la marca Red Bull, con el fin de conocer el comportamiento de la empresa en la red y cómo aplica las prácticas de marketing en este entorno.**
- **Profundizar en el estudio de la Minería de Opiniones y Análisis de Sentimientos a través de algoritmos automáticos, con base a las investigaciones existentes como por ejemplo (Okazaki, Díaz-Martín, Rozano, & Meneñdez-Benito, 2014), con el fin de detectar y aplicar las técnicas y herramientas más eficientes y que proporcionen una mayor fiabilidad de resultados.**



- Finalmente se recomienda un estudio basado en la co-ocurrencia de palabras, identificando las palabras que están relacionadas entre sí, según el grado de proximidad y frecuencia que aparecen en los tweets, plasmándolos en forma de grafos 3D.

## 8. Referencias Bibliográficas

---

Ahmad, K. (2011). Affective computing and sentiment analysis. In K. Ahmad (Ed.), *Affective computing and sentiment analysis* (pp. 1-150) Springer.

Ampofo, L. (2011). The social life of real-time social media monitoring. *Participations*, 8(1), 21-47.

Bello, G., Menéndez, H., Okazaki, S., & Camacho, D. (2013). Extracting collective trends from twitter using social-based data mining. *Computational collective intelligence. technologies and applications* (pp. 622-630) Springer.

Berners-Lee, T., Fischetti, M., & Foreword By-Dertouzos, M. L. (2000). *Weaving the web: The original design and ultimate destiny of the world wide web by its inventor* HarperInformation.

Bhuta, S., & Doshi, U. (2014). A review of techniques for sentiment analysis of twitter data. *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference On*, pp. 583-591.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.

Breen, J. O. (2012). Mining twitter for airline consumer sentiment. *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*, , 133.

Chen, H., & Zimbra, D. (2010). AI and opinion mining. *Intelligent Systems, IEEE*, 25(3), 74-80.

Colaboradores de Wikipedia. (2014). Web 2.0. Recuperado, 21-06-2014, 09:53 UTC, 2014, de [http://es.wikipedia.org/w/index.php?title=Web\\_2.0&oldid=76874059](http://es.wikipedia.org/w/index.php?title=Web_2.0&oldid=76874059)



Cormode, G., & Krishnamurthy, B. (2008). Key differences between web 1.0 and web 2.0. *First Monday*, 13(6)

Davidov, D., Tsur, O., & Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 241-249.

Esuli, A., & Sebastiani, F. (2010). Machines that learn how to code open-ended survey data. *International Journal of Market Research*, 52(6), 775-800.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070), 900-901.

Hassan, M., Mohamed, Ameen, M., Mohamed, Hamed, M., Abdelkader, & Mahmoud, M., Mohamed. (2013). Sentiment classification for product reviews, Unpublished manuscript.

Hodes, R. L., Cook, E. W., & Lang, P. J. (1985). Individual differences in autonomic response: Conditioned association or conditioned fear? *Psychophysiology*, 22(5), 545-560.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168-177.

Hunt, T. (2010). *O poder das redes sociais* Editora Gente Liv e Edit Ltd.

The logo for 'cim' consists of a stylized graphic of three interconnected circles above the lowercase letters 'cim'.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, 53(1), 59-68.

Kotler, P., & Armstrong, G. (2008). *Princípios de marketing* Pearson Educación.

Kozinets, R. V. (1998). On netnography: Initial reflections on consumer research investigations of cyberculture. *Advances in Consumer Research*, 25(1), 366-371.

Kozinets, R. V. (2010). *Netnography: Doing ethnographic research online* Sage Publications.

Lang, P. J. (1980). *Behavioral treatment and bio-behavioral assessment: Computer applications*.



Larose, D. T. (2005). In John W. (Ed.), *Discovering knowledge in data : An introduction to data mining* [Discovering knowledge in data : an introduction to data mining ] (1st ed.). United States of America:

Lévy, P. (2004). *Inteligencia colectiva. Por Una Antropología Del Ciberespacio*,

Lévy, P. (Ed.). (2001). *Cyberculture* (Robert Bononno Trans.). (1st ed.). Francia:

Li, Y., & Li, T. (2011). *Deriving marketing intelligence over microblogs*. *System Sciences (HICSS)*, 2011 44th Hawaii International Conference On, pp. 1-10.

Liu, B., Hu, M., & Cheng, J. (2005). *Opinion observer: Analyzing and comparing opinions on the web*. *Proceedings of the 14th International Conference on World Wide Web*, pp. 342-351.

Liu, B. (2007). *Web data mining: Exploring hyperlinks, contents, and usage data* Springer-Verlag.

Maguregui, C. (2009). *Wikipedia, enciclopedia británica, redes, inteligencia colectiva y trabajo colaborativo*. Recuperado, 19-07-2014, de <http://portal.educ.ar/debates/sociedad/wikipedia-enciclopedia-britani.php>

Moore, M. (2009). *Encyclopaedia britannica fights back against wikipedia*. Recuperado, 22-07-2014, de <http://www.telegraph.co.uk/culture/books/booknews/4318176/Encyclopaedia-Britannica-fights-back-against-Wikipedia.html>

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How japanese companies create the dynamics of innovation* Oxford university press.

Okazaki, S., Díaz-Martín, A., M., Rozano, M., & Meneñdez-Benito, H., D. (2014). *How to mine brand tweets: Procedural guidelines and pretest* *International Journal of Market Research*, , 1-10.

Onishi, H., & Manchanda, P. (2012). *Marketing activity, blogging and sales*. *International Journal of Research in Marketing*, 29(3), 221-234.

O'Reilly, T. (2007). *What is web 2.0: Design patterns and business models for the next generation of software*. *Communications and Strategies*, 65(1), 17-37.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

Polo, J. D. (2009). *Twitter: Para quien no usa twitter* Bubok.

R Foundation. (2002). What is R? Introduction to R. Recuperado, 21-06-2014, de <http://www.r-project.org/>

Reyes, A., Rosso, P., & Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74, 1-12.

Rheingold, H. (1996). *La comunidad virtual: Una sociedad sin fronteras* Gedisa Barcelona.

Rodríguez, R., A., & Santamaría, C., P. (2012). Análisis del uso de las redes sociales en internet: Facebook y twitter en las universidades españolas. *Revista ICONO14. Revista Científica De Comunicación Y Tecnologías Emergentes*, 10(2), 228-246.

Russell, M. A. (2013). *Mining the social web: Data mining facebook, twitter, LinkedIn, google , GitHub, and more* " O'Reilly Media, Inc."

Sagolla, D. (2009). How twitter was born. Recuperado, 16-06-2014, de <http://www.140characters.com/2009/01/30/how-twitter-was-born/>

cim

Sameh, A. (2013). Using social networks in political elections. The 27th Annual Conference of the Japanese Society for Artificial Intelligence, Japon. , 27th.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (in litt.). *Sentiment analysis: A review and comparative analysis of web services.*

Solo Marketing. (2014). *Sentiment analysis de la campaña de campofrío.* Recuperado, 19-06-2014, de <http://www.solomarketing.es/sentiment-analysis-de-la-campana-de-campofrio/>

Tenzer, S., M., Ferro, O., & Palacios, N. (2009). *Redes sociales virtuales: Personas, sociedad y empresa.* Cátedra Introducción a La Computación, FCEA,

Toffler, A. (1980). In Bantam Books (Ed.), *THE THIRD WAVE [THE THIRD WAVE]*. United States and Canada:



Tolsan, V. (2012). Ejemplos de seguimiento de campañas políticas en las redes sociales. Recuperado, 20-06-2014, de <https://blog.websays.com/como-detectar-la-opinion-publica-de-una-campana-electoral-gracias-a-las-redes-sociales/>

Turban, E. K., McKay, D., Marshall, J., Lee, P., & Viehland, J. (2008). Electronic commerce A managerial perspective.

Twitter, I. (2013). Política de protección de datos de twitter Recuperado, 17-07-2014, de <https://twitter.com/privacy>

Twitter, I. (2014a). About twitter. Recuperado, 17-07-2014, de <https://about.twitter.com/company>

Twitter, I. (2014b). Cómo twitter puede ayudar a su negocio. Recuperado, 17-07-2014, de <https://business.twitter.com/es/how-twitter-can-help-your-business>

Twitter, I. (2014c). Desigual ¿Cómo consigue una marca de moda maximizar el alcance y la interacción en su campaña de año nuevo?. Recuperado, 17-07-2014, de <https://business.twitter.com/es/success-stories/desigual>

Twitter, I. (2014d). ESPN ¿De qué manera una red de deportes impulsa más tráfico a sus artículos y videos? Recuperado, 17-07-2014, de <https://business.twitter.com/es/success-stories/espn>

Twitter, I. (2014e). Historias de éxito: Gran empresa. Recuperado, 17-07-2014, de <https://business.twitter.com/es/success-stories/company-size/xl-biz>

Twitter, I. (2014f). ING direct españa. Recuperado, 17-07-2014, de <https://twitter.com/INGDIRECTes>

Twitter, I. (2014g). The story of a tweet. Recuperado, 17-07-2014, de <https://about.twitter.com/what-is-twitter/story-of-a-tweet>

cim

