

INVESTIGACIÓN

Recibido: 16/03/2021 --- Aceptado: 16/03/2021 --- Publicado: 24/05/2021

ANÁLISIS DE LOS PROCESOS DE TRATAMIENTO DE INFORMACIÓN EN UN ESTUDIO DE ANÁLISIS DE SENTIMIENTO UTILIZANDO LA TECNOLOGÍA DE GOOGLE

Analysis processes treatment study information in sentiment analysis using technology Google

  Ángel Quintana Gómez¹: Universidad del Atlántico Medio. España.
angel.quintana@pdi.atlanticomedio.es

Cómo citar el artículo:

Quintana Gómez, A. (2021). Análisis de los procesos de tratamiento de información en un estudio de análisis de sentimiento utilizando la tecnología de Google. *Vivat Academia. Revista de Comunicación*, 154, 41-55.
<http://doi.org/10.15178/va.2021.154.e1336>
<http://www.vivatacademia.net/index.php/vivat/article/view/1336>

RESUMEN

En los últimos años el *Big Data* se abre camino entre las principales herramientas de análisis de mercado, vinculándose a las técnicas de *Machine Learning* con el fin de aprender sobre los datos que se posee. Una de las áreas de mayor crecimiento es el Procesamiento de Lenguaje Natural que proporciona al investigador datos sobre estructuras y significados de texto. Con el fin de profundizar en esta área, *Google* ha creado la *API Natural Language*, permitiendo a los investigadores trabajar con distintos aspectos de las funciones del lenguaje entre ellas el análisis de sentimiento, proporcionando información sobre la opinión emocional predominante de un contenido seleccionado previamente y permitiendo obtener un score que analiza la valencia de las emociones con valores dicotómicos. El objeto de este estudio es analizar los distintos procesos que un investigador tiene que utilizar para obtener información útil para sus investigaciones. Desde la extracción de información, hasta la obtención de datos que ayuden al investigador a obtener conclusiones, se desarrolla un largo proceso del tratamiento de la información. El estudio nos mostrará como las diversas herramientas de las que dispone *Google* en su plataforma *Google Cloud Platform*, aportan a un investigador el apoyo necesario para el desarrollo de su trabajo, una vez que ya se cuenta con la información a analizar. Además, se complementará con herramientas

¹ Ángel Quintana Gómez: Doctor en Comunicación Audiovisual, Publicidad y Relaciones Públicas.

de rastreo para la extracción del texto que se desea, en función de donde se encuentre esta.

PALABRAS CLAVES: API Natural Language- Análisis de sentimiento-Big Query-Crawler- Google Cloud Platform- Google Data Prep- Google Data Studio- Twitter

ABSTRACT

In recent years, Big Data has made its way amongst the main market analysis tools, linking itself to machine learning techniques in order to learn about the data owned. One of the fastest growing areas is natural language processing, which provides the researcher with data on text structures and meanings. In order to deep in into this area, Google has created the natural language API, allowing researchers to work with different aspects of language functions, including sentiment analysis, providing information on the predominant emotional response to a previously selected content, and allowing it to obtain a score that analyzes the valence of emotions with dichotomous values. The object of this study is to analyze the different processes that a researcher has to use to obtain useful information for their research. From the extraction of information to obtaining data that helps the researcher to draw conclusions, a long process of information processing is developed. The study will show us how the various tools available to Google on its own Google Cloud Platform provide a researcher with the necessary support for the development of their work, once the information to be analyzed is already available. In addition, it will be complemented with tracking tools to extract the desired text, depending on where it is.

KEYWORDS: API Natural Language- Big Query; Crawler- Google Cloud Platform- Google Data Prep- Google Data Studio- Sentiment Analyses- Twitter

1. INTRODUCCIÓN

Existe un constante crecimiento de investigaciones relacionadas con el neuromarketing, desde que las nuevas técnicas de neuroimagen de las ciencias sociales han supuesto un giro en la investigación social (Alcover de la Hera, 2008; Harmon-Jones & Beer Jennifer, 2009; Ramachandran, 2011). Esto es debido a la relevancia que va adquiriendo el mundo de las emociones en el estudio de los consumidores y a que las metodologías tradicionales basadas en las ciencias sociales no dan las respuestas correctas; no tanto por el deseo de los sujetos de no mostrar la realidad, sino por desconocerla, el neuromarketing puede acertar con las respuestas idóneas (Martínez Herrador et al., 2020). Y al igual que ocurre en otras áreas de investigación, se obtienen mejores conclusiones en función del volumen de datos que se utiliza. Cuando se trata en una investigación con un volumen relevante de datos, es preciso trabajar con técnicas de *Big Data*, ya que estos nos permitirán procesar una gran cantidad de ellos, de una forma muy rápida.

En función de las investigaciones, nos podemos encontrar con un volumen grande de datos escritos, o bien de imágenes. El área del *Big Data* nos va a permitir trabajar ambas situaciones. Estudios como el de Boubela (2016) muestra la aplicación de técnicas de *Big Data*, en concreto a través de Apache Park, en los que analiza datos de resonancia magnéticas. El desarrollo del *Big Data* se optimiza no solo para ser capaz de trabajar con el fin de agilizar los procesos y que estos sean más rápidos, sino también para realizar análisis más complejos (Boubela et al., 2016; Eklund et al., 2012).

En el área del neuromarketing, y relacionado con el Análisis de Sentimiento, se ha utilizado *Hadoop* para analizar un gran volumen de datos (Singh et al., 2019). La intrusión del *Big Data* en esta área, según Bayod (2018), aplicándolo al área de la política, se convierte en una herramienta tecnológica que permite personalizar los mensajes, que analiza las actitudes, y sentimientos de los electores, con la información, tanto estructurada como no estructurada, obtenida de las redes sociales. Estudios como el de Hernández-Fernández (2019) combinan tanto técnicas de reconocimiento facial como análisis de sentimientos, para analizar y comprender la valoración del lanzamiento de *Google Glass* en los consumidores. Aunque con el desarrollo de investigaciones, se han ido optimizando todo lo referente a la minería de opinión, bien es cierto, que inicialmente presentaba problemas de detección de la correcta polaridad debido a aspectos como la ironía o el sarcasmo; así como el tratamiento de las negaciones o simplemente detectar el Spam (Mart et al., 2014).

A la hora de llevar a cabo dichas investigaciones, es bastante común depender de un equipo especialista que configure y gestione toda la programación para la investigación; qué según que casos, puede llevar desde la extracción de los datos, su estructuración y limpieza, para su posterior almacenamiento y representación.

Pero con el paso de los años, y los desarrollos tecnológicos en la nube, hay empresas que está ofertando servicios, que permiten bien a las organizaciones o investigadores, utilizar su software sin necesidad de programación y a unos costes bajos, posibilitando e impulsando el investigar con el tratamiento de gran cantidad de datos. Actualmente se denominan, soluciones de *Cloud Computing*. Según Rani (2014), las principales características a nivel global son: mantenimiento más económico, interfaz que facilita la interacción entre usuario y ordenador, configuración en función de las necesidades, sin necesidad de actualización, y tiene una plataforma de gran escala permitiendo la ejecución de muchas aplicaciones a la vez. Por tanto, en la comunidad científica, se puede encontrar con investigaciones que han utilizado sus investigaciones en la nube, en concreto en *Platform as a Service* (PaaS)

Se pueden desarrollar en 3 grandes áreas: *Infrastructure as a Service* (IaaS), *Platform as a Service* (PaaS) y *Software as a Service* (SaaS).

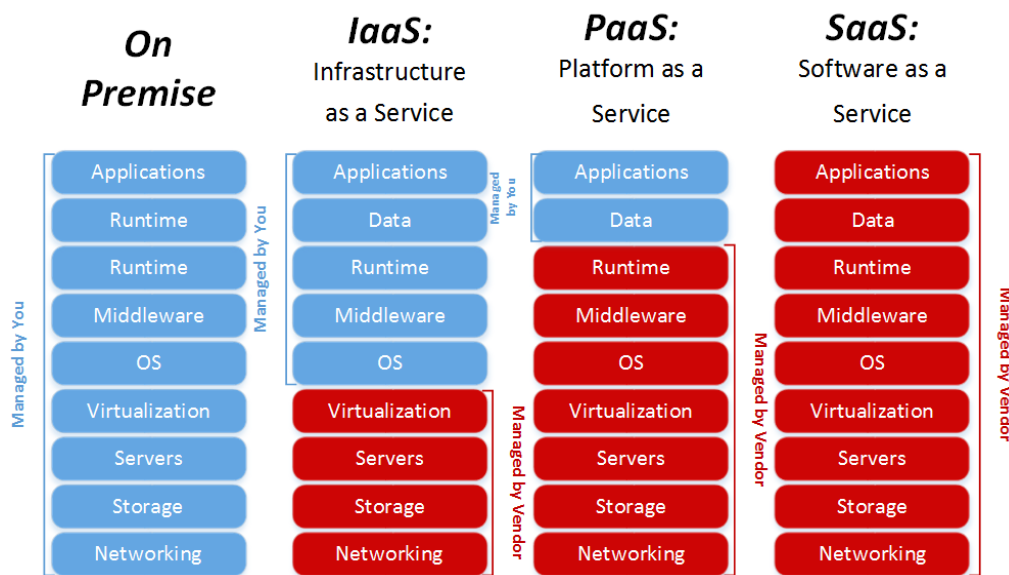


Figura 1. Tipos de plataformas en la nube.
Fuente: crucial.com.au (2013)

Aunque existen investigaciones que tratan de crear nuevas plataformas para trabajar áreas específicas (Liu *et al.*, 2019), otros estudios demuestran el éxito de las ya existentes en el mercado, validando los resultados a través de las aplicaciones utilizadas (Agavanakis *et al.*, 2019; Quintana, 2019).

En la investigación *The state of SQL-on-Hadoop in the Cloud* (Poggi *et al.*, 2016), se lleva a cabo un comparativo entre los servicios PaaS de *Hadoop* con los 4 principales proveedores en la nube: *Amazon*, *Azure*, *Google* y *Rackspace*, en el que concluye que todos ellos pueden mejorar en algunos aspectos. Pero consideró a *Google* la PaaS con facturación más atractiva. *Google Cloud Platform*, es considerada una de los líderes entre las APIs en la nube, aunque cuenta con el inconveniente de tener la documentación de la misma presenta notables inconvenientes que dificultan su implementación (Challita *et al.*, 2018).

En una reciente publicación, *Forrester* nombra a *Google Cloud* uno de los líderes en el informe “*The Forrester Wave: Data Management for Analytics, Q1 2020*” (Yuhanna, 2020).

2. OBJETIVOS

La presente investigación, tiene por objetivo mostrar los distintos pasos para trabajar en una investigación de neuromarketing con *Big Data*. En concreto, con el entorno de *Google Cloud Platform*, que cuenta con más de 100 aplicaciones a disposición de cualquier investigador. La investigación, se centra en el análisis de sentimiento de tuits donde se nombra la palabra *neuromarketing* en España, durante diciembre de 2020. Desde su extracción, hasta su representación, describiendo el tipo de perfil de los usuarios que generan dicho contenido en función cuando y qué tuita.

3. METODOLOGÍA

Se utilizará como plataforma de investigación *Google Cloud Platform*. Se trata de una plataforma en la nube que soporta más de 100 aplicaciones diferentes. Las áreas que trabaja se pueden agrupar en aprendizaje automático, infraestructura, plataforma y aceleradores, gestión de APIs, computación, contenedores, analíticas de datos, bases de datos, herramientas de desarrollo, entornos híbridos y multinubes, internet de las cosas, herramientas de gestión, contenido multimedia y videojuegos, migración, redes, operaciones, seguridad e identidad y computación sin servidor.

Como fuente de información, se utilizará un *Data Set* de la red social de *Twitter*. En concreto un rastreo la palabra *neuromarketing*, incluyendo *hashtags*. Para ejecutar el rastreo, se debe vincular *Google Cloud Platform* con *Twitter*. Para ello, se utilizará la propia API de la red social, garantizando de esta forma la calidad del contenido registrado. Una de las características de las APIs, es que tienen una ventana temporal determinada, para esta investigación se cuenta con una ventana temporal de 9 días. Por lo que manualmente se realiza esta acción.

Para el tratamiento de la información, se utilizan las siguientes aplicaciones que se encuentran en la plataforma: *Google Cloud Storage*, *Google DataPrep*, *Google Big Query*, *Google Data Studio* y *Natural Language* de *Google*.

Google Cloud Storage es un almacén de datos ubicado en la nube. Permite cuatro tipos diferentes de almacenamiento, en función de la disponibilidad que se desea de los datos. Aportando así, una reducción en los costes. Los almacenamientos disponibles son *Standard*, *Nearline*, *Coldline*, y *Archive*.

Google DataPrep es un servicio inteligente de datos en la nube, que permite examinar, limpiar y preparar datos estructurados para crear modelos de aprendizaje automático. Una de las ventajas de este servicio es que con cualquier interacción que se realice con la interfaz, automáticamente predice y sugiere diversas transformaciones que pueden llevarse a cabo.

Google BigQuery es un almacén de datos multinube que permite realizar consultas interactivas de grandes conjuntos de información, a través de *Structured Query Language* (SQL). Permite la programación de *Queries* y con su servicio *BigQuery Data Transfer Service* permite la transferencia automática de datos a fuentes externas.

Google Data Studio es una herramienta que convierte los datos en paneles e informes. Puede conectarse a los productos de *Google Marketing Platform*, a productos de consumo de *Google*, a bases de datos como *BigQuery*, así como *Google Cloud Storage*.

Natural Language de *Google* muestra la estructura y significado de los textos. Está preparado para extraer información relacionada con personas, lugares o eventos. Con la API de *Natural Language* y debido a su entrenamiento previo automatizado, la

herramienta realiza análisis de sentimientos, aportando información sobre si los textos analizados son positivos, negativos o neutros. Esta información la proporciona con un índice entre -1 y +1.

A continuación, se detallan las fases de la investigación llevadas a cabo en el ecosistema de Google:

El rastreo de la información es la primera fase de la investigación. Se crea un csv que incluya las palabras claves a rastrear. Dicho csv, se sube y almacena en la plataforma *Google Cloud Storage* que, a su vez, ha sido conectada a la API de *Twitter*; y donde con una frecuencia de 9 días, se ejecuta el rastreo de la información contenida en el csv. El resultado que se obtiene se almacena en una tabla en *Google BigQuery*. Debido a que hay que ejecutar varios análisis, en las sucesivas ejecuciones, se añadirá al contenido a la ya existente tabla; evitando así, la creación de múltiples tablas ya que supondría tener que ejecutar diversas queries para poder unificarlas.

Debido a que la investigación es de 30 días, la acción manual se realiza en cuatro ocasiones. Esto supone duplicación de datos, ya que el rastreo del día 30 duplicará aquellos tuits registrados entre el día 22 y 27 que se realiza el mismo día 27 de diciembre. Los datos que se obtienen de cada tuit son: fecha de creación, id del tuit, texto, usuario, URL, retuits, favoritos, seguidores y si el emisor del tuit es una cuenta verificada.

El tratamiento y limpieza de la información es el siguiente paso, desde *Google DataPrep* se importa la tabla de *Google BigQuery*. *Google DataPrep*, permite examinar, limpiar y preparar datos estructurados para crear modelos de aprendizaje automático; y a través de los *Recipes* del software se crean recetas donde se aplicarán una serie de normas que hay que configurar. Entre ellas, las más relevantes son:

1. Identificar aquellos tuits donde aparece la keyword *neuromarketing*. Existe la posibilidad de que durante la extracción de la información entren tuits que no cumplen en su totalidad con las exigencias requeridas.
2. Eliminar duplicidades. Al realizar un último rastreo que solapaba días de rastreo con la ejecución anterior, existirán tuits idénticos, que deben ser eliminados. Y en ocasiones, es posible que un mismo tuit se inserte en la tabla varias veces.
3. Eliminar tuits fuera de la fecha de análisis, ciñéndonos así únicamente al mes de diciembre de 2020

Estas primeras recetas, nos aportan un contenido limpio y sin errores, pero también podemos aplicar otras recetas que nos permitan extraer información útil para la investigación. Son las siguientes:

1. Creación de una columna donde aparezca el primer *hashtag* utilizado en el tuit. Igualmente, otra columna donde figure el segundo *hashtag* utilizado, y así hasta un máximo de 10. Permitiendo conocer con que *hashtags* se asocian los tuis donde se mencionas la keyword a analizar.

2. Creación de una columna que identifique los tuits que han utilizado #neuromarketing en su tuit, independientemente de que se haya escrito en mayúscula o minúscula.

Esta tecnología permite actualizar manualmente el archivo que procede de *BigQuery*, por lo que, si se decide prolongar la investigación un mes más, al subir nuevamente el archivo se aplicarían las recetas a toda la nueva información subida.

La tercera fase de la investigación consiste en migrar esta nueva tabla de información a *Google BigQuery*. Ya que, además de estar saneada, es más rica y contiene nueva información como los *hashtag*, que antes no existía. Este proceso se puede realizar manualmente o se puede automatizar. Por lo que *Google BigQuery* permite automatizar la descarga de la información tanto a *Google Storage* como a *Google BigQuery*. Todo dependerá de las necesidades del proyecto.

Debido a que la siguiente fase del proceso, realizar el análisis de sentimiento de los textos de los tuits, implica que la *API Natural Language* de *Google* se nutre de las tablas de *Google BigQuery*, ambas aplicaciones deben estar vinculadas. La ejecución de la *API Natural Language* de *Google*, ofrece como resultado una nueva tabla que tan solo contiene las siguientes columnas: URL, texto, *score*, *magnitud*, frases, *score* de frases y *magnitud* de frases. Automáticamente, se genera una nueva tabla en *Google BigQuery* con dicha información.

Al igual que con la información extraída de *Twitter*, es recomendable tratar la información que proveen los resultados de la *API Natural Language*. Por un lado, para unir los contenidos de las tablas, la previa obtenida del rastreo de *Twitter* y que ya está limpia junto con el resultado del análisis de sentimiento. Es cierto, que dicha vinculación es posible realizarla con una simple *query* de *union*, pero lo cierto es que es posible modificar visualmente la información extraída de la API para su posterior visualización en gráficos y poder lograr un entendimiento más rápido de la información. Para realizar esta acción, retomamos la acción de importar la tabla desde *Google BigQuery* a *Google DataPrep*. Las recetas que utilizaremos son las siguientes:

1. Unión de las tablas ya mencionadas.
2. Redondear a dos decimales los datos de *score* y *magnitud*; ya que la herramienta llega a proveer hasta de 10 decimales en algunos casos, no en todos.
3. Creación de una fórmula para convertir la puntuación del *score* en una etiqueta de texto según sea positivo, negativo o neutro.

CASE ([score < -0.25,'negativo', score > 0.25,'positivo','neutro'])

Con el tratamiento realizado a las tablas, de tener 2 tablas, donde una está formada por 12 columnas y otra por 8, teniendo dos de ellas iguales (url y texto), obtenemos una única tabla que consta de 29 columnas. A su vez, de contar inicialmente con un total de 482 filas, contaremos únicamente con 429.

Y último paso, previo al análisis de la información, es la visualización de esta. La herramienta de *Google Data Studio* permite la vinculación con las tablas de *BigQuery*, de forma que se pueden crear gráficos y tablas personalizados. El usuario dispone de todas las columnas de las tablas, para crear nuevas de forma independiente utilizando únicamente los datos de las columnas que son de interés. Además, cuenta con la opción de filtro sobre el contenido de las tablas; de forma, que podemos realizar un gráfico con la información de aquellos usuarios que tengan más de 100 seguidores, o la de aquellos que han utilizado un determinado *hashtag*.

4. DISCUSIÓN

El primer dato que se obtiene de la investigación es el resultado que el *Crawler* vuelca sobre *Google BigQuery*, donde aparecen los resultados de la aparición de la palabra *neuromarketing* en *Twitter*, un total de 489 tuits. Estos son vinculados a *Google DataPrep*, donde son limpiados y se eliminan duplicidades y posibles tuits que no cumplan con las condiciones de la investigación. Finalmente, el resultado donde se tiene en cuenta al cien por cien de los usuarios que han nombrado la palabra *neuromarketing*, independientemente que haya sido con mayúscula, minúscula o con *hashtag*, durante el mes de diciembre de 2020, es de un total de 429 tuits que han sido tuiteados por 209 usuarios, suponiendo una media de 2,05 tuits por cuenta de *Twitter*. Y si se realiza un sumatorio de los valores obtenido a nivel de score de cada uno de los tuits, obtenemos una puntuación total de 113,6.

Tras realizar el análisis de sentimiento a los 429 tuits, obtenemos que 197 obtienen un score superior a 0.25, por lo que se consideran positivos; y el total de estos tuits positivos obtienen un score medio de 0,42. La difusión de estos mensajes, fueron difundidos por 106 usuarios distintos y el valor total de la polaridad asciende a 83,7.



El total de las menciones neutras, lo constituyen 320 tuits, que fueron difundidos 120 usuarios, obteniendo un score medio de 0,13 y sumando un total de 30,5. El sumatorio de los scores individuales, tanto los que están en la franja positiva como negativa (de -0,25 a 0,25), resultan alcanzar la cifra de score total de 30,5

Respecto a las menciones negativas, únicamente se realizaron 2 que fueron lanzadas por cuentas diferentes. Su score medio es de 0,3 alcanzando un valor total de -0,6.

Tabla 1. Resultados Análisis de Sentimiento

<i>Tipo Tuits</i>	<i>Nº Usuario</i>	<i>Nº Tuits</i>	<i>Score Medio</i>	<i>Sumatorio Score</i>
Todos	209	429	0,26	113,6
Positivos	106	197	0,42	83,7
Negativos	2	2	-0,3	-0,6
Neutros	120	230	0,13	30,5

Fuente: Elaboración propia

Analizando la difusión de los tuits obtenemos una media diaria de 13,8 tuits. El día con mayor publicación de tuits es el 15 de diciembre, con 28 tuits y un score medio superior a la media, 0,29. Los 28 tuits están constituidos por 20 positivos, 7 neutros y 1 negativo. Incluía, además, 12 retuits y fueron emitidos por 25 usuarios distintos; no siendo ninguna de ellas una cuenta verificada. El contenido del tuit más retuiteado ese día es el siguiente: "35 estrategias de #Neuromarketing para crear títulos IRRESISTIBLES   <https://t.co/ATJhFJ1dHF> por @SalimaSanchez vía @FerRubioA #Marketing #marketingcontenidos #MarketingDigital <https://t.co/GyAztawsWa>"

El día con menor difusión corresponde al 6 de diciembre, día festivo que se celebra la Constitución española. Tan solo se publicó un tuit. Salvo el fin de semana del 19 y 20 de diciembre, que hubo bastante actividad, los fines de semana desciende el número de publicaciones. El 9 de diciembre, es la fecha donde se alcanza un score medio más positivo, siendo de 0,45 con un total de 13 publicaciones, dos de ellas alcanzaron un score de 0,8 y 0,9; ninguno fue retuiteado.

Tabla 2. Distribución de tuits por día

<i>Tipo Tuits</i>	<i>Nº Tuits</i>	<i>Score Medio</i>	<i>Sumatorio Score</i>
1/12/20	10	0,28	2,80
2/12/20	13	0,25	3,30
3/12/20	23	0,25	5,70
4/12/20	25	0,32	8,00
5/12/20	23	0,20	4,50
6/12/20	1	0,30	0,30
7/12/20	24	0,25	5,90
8/12/20	15	0,22	3,30
9/12/20	13	0,45	5,90
10/12/20	12	0,24	2,90
11/12/20	11	0,34	3,70
12/12/20	5	0,22	1,10
13/12/20	6	0,28	1,70
14/12/20	12	0,24	2,90
15/12/20	28	0,29	8,20
16/12/20	16	0,29	4,60
17/12/20	2	0,15	0,30
18/12/20	14	0,38	5,30
19/12/20	13	0,33	4,30
20/12/20	20	0,26	5,10
21/12/20	23	0,24	5,60
22/12/20	21	0,25	5,30

Quintana Gómez, A.
Análisis de los procesos de tratamiento de información en un estudio de análisis de sentimiento utilizando la tecnología de Google

23/12/20	19	0,29	5,60
24/12/20	14	0,27	3,80
25/12/20	10	0,18	1,80
26/12/20	3	0,20	0,60
27/12/20	7	0,21	1,50
28/12/20	19	0,17	3,20
29/12/20	10	0,34	3,40
30/12/20	10	0,14	1,40
31/12/20	7	0,23	1,60

Fuente: Elaboración propia.

Respecto a la distribución por franja horaria, la mayor actividad se concentra desde las 6:00 AM hasta las 17:00 PM. Hay que destacar los picos de difusión que se producen durante la mañana entre las 9:00 y 10:00, y las 11:00 y 12:00; así como la que se realiza entre las 22:00 y 23:00 horas. Las horas con menos difusión y que por lo tanto su sumatorio de score no es relevante, se produce entre las 00:00 de la noche y las 6:00 de la mañana. A lo largo del mes analizado, no existe ninguna hora completa donde no se haya publicado al menos un tuit.

Tabla 3. Distribución de tuits por hora

<i>Hora</i>	<i>Nº Tuits</i>	<i>Score Medio</i>	<i>Sumatorio Score</i>
0:00	2	0,45	0,90
1:00	1	0,20	0,20
2:00	1	0,20	0,20
3:00	1	0,10	0,10
4:00	2	0,20	0,40
5:00	6	0,20	1,20
6:00	18	0,25	4,50
7:00	34	0,26	8,90
8:00	29	0,19	5,60
9:00	40	0,24	9,70
10:00	23	0,37	8,60
11:00	44	0,29	12,90
12:00	23	0,29	6,70
13:00	24	0,21	5,10
14:00	30	0,31	9,20
15:00	25	0,29	7,30
16:00	34	0,30	10,30
17:00	22	0,26	5,80
18:00	14	0,24	3,30
19:00	11	0,20	2,20

Quintana Gómez, A.
Análisis de los procesos de tratamiento de información en un estudio de análisis de sentimiento utilizando la tecnología de Google

20:00	8	0,15	1,20
21:00	11	0,15	1,60
22:00	21	0,29	6,00
23:00	5	0,34	1,70

Fuente: Elaboración propia

El total de los 429 tuits publicados, se conforman únicamente por 295 tuits; ya que 134 de ellos corresponden a retuits de las publicaciones originales, representan el 31,24% del total. El score medio de los tuits retuiteados es inferior a la media, siendo de 0,24 frente a 0,26. Por el contrario, los tuits no retuiteados forman el 68,76% y tienen un score medio de 0,28.

Tabla 4. Retuits

<i>Tipo Tuits</i>	<i>Nº Tuits</i>	<i>Porcentaje</i>	<i>Score Medio</i>
Todos	429	100,00%	0,26
Retuiteados	134	31,24%	0,24
No retuiteados	295	68,76%	0,28

Fuente: Elaboración propia

Analizando la inclusión del *hashtag* #neuromarketing en los tuits, encontramos que en el 66,2% de ellos está presente, un total de 284 tuits. El análisis de sentimiento muestra que el 51,1% son considerados positivos, el 48,2% neutro y también este *hashtag* está incluido en los dos negativos que se difundieron durante el periodo analizado.

Tabla 5. Tuits que incluyen el *hashtag* #neuromarketing

<i>Tipo Tuits</i>	<i>Nº Tuits</i>	<i>Porcentaje</i>	<i>Score Medio</i>
Todos	429	100,00%	0,26
Con <i>hashtag</i>	284	66,20%	0,29
Sin <i>hashtag</i>	145	33,80%	0,21

Fuente: Elaboración propia

Los tuits retuiteados que incluyen este *hashtag* cuentan con un leve score medio superior a la media, siendo de 0,27 frente a 0,26. El 69,01% que no fue retuiteado, alcanzan un score medio superior de 0,3.

Tabla 6. Retuits que incluyen el *hashtag* #neuromarketing

Quintana Gómez, A.
Análisis de los procesos de tratamiento de información en un estudio de análisis de sentimiento utilizando la tecnología de Google

<i>Tipo Tuits</i>	<i>Nº Tuits</i>	<i>Porcentaje</i>	<i>Score Medio</i>
Todos	284	100,00%	0,26
Retuiteados	88	30,99%	0,27
No retuiteados	196	69,01%	0,3

Fuente: Elaboración propia.

También se han analizado las cuentas verificadas que incluyeron en sus tuis la palabra Neuromarketing, encontrando únicamente 2 usuarios. Estas dos cuentas de *twitter* corresponden a Universidad UDIMA (@UDIMA), que cuenta con más de 15.400 seguidores; y el Ayuntamiento de Valladolid (@AyuntamientoVLL) con más de 49.400 seguidores. Cada uno de ellos, publicó únicamente un tuit, ambos neutros y alcanzaron un score medio de 0,2.

Tabla 7. Tuits de usuarios verificados

<i>Tipo Tuits</i>	<i>Nº Tuits</i>	<i>Porcentaje</i>	<i>Score Medio</i>
Todos	429	100,00%	0,26
Verificados	2	0,47%	0,2
No verificados	427	99,53%	0,27

Fuente: Elaboración propia.

Analizando el número de seguidores de las cuentas de los tuits detectamos que el 27,5% de los tuits, han sido realizados por cuentas que cuentan con menos de 500 seguidores; aunque cabe destacar que aquellos que tienen entre 201 y 500 seguidores difunden, por encima de la media, tuits positivos. Por el contrario, aquellas cuentas que tienen entre 501 y 10.000 seguidores, lanzan tuits con un score medio inferior a la media; y suponen más del 50% de los tuits publicados. Finalmente, los tuits lanzados cuyas cuentan tienen más de 10.000 seguidores y que representan únicamente al 12,59% de los tuits lanzados, obtienen el score medio más alto, un 0,37.

Tabla 8. Resultados en función de número de seguidores

<i>Tipo Tuits</i>	<i>Nº Tuits</i>	<i>Porcentaje</i>	<i>Score Medio</i>
Todos	429	100,00%	0,26
De 1 a 200	53	12,35%	0,26
De 201 a 500	65	15,15%	0,32
De 501 a 1.000	102	23,78%	0,23
De 1001 a 10.000	155	36,13%	0,23
Más de 10.000	54	12,59%	0,37

Fuente: Elaboración propia

5. DISCUSIÓN

Tras el análisis realizado, podemos concluir como la metodología aplicada de plataforma en la nube, permite elaborar trabajos de investigación sin necesidad de tener conocimientos profundos a nivel de programación. Bien es cierto, que el ecosistema hay que crearlo, pero una vez esté operativo, el investigador contará con la independencia necesaria para poder realizar las investigaciones que considere y a un bajo coste. Además, dicho sistema servirá para utilizarlo para otras investigaciones.

Para el ahorro de tiempo, es fundamental el diseño de la investigación, y analizar con detalle las posibilidades que tiene cada una de las APIs que se van a utilizar, ya que a través de la automatización de procesos, tanto en *Google BigQuery* como en *Google DataPrep*, podremos evitar la reiteración sistemática de la misma operación. Es importante tener en cuenta que la actualización de las APIs son continuas, por lo que es recomendable consultarlas por si sacan nuevas funcionalidades que ayuden a facilitar o profundizar la investigación.

Para obtener unos resultados válidos en la investigación, es esencial un buen filtrado de la información extraída. Como se observa en la investigación, se ha trabajado con 429 tuits, pero se obtuvieron de la extracción inicial un total de 482 tuits, un 12,35% más de valores duplicados o erróneos. En el caso de realizar un análisis de alguna otra palabra con más volumen de difusión en Twitter, hubiera llevado exactamente el mismo tiempo llevarlo a cabo, la única diferencia hubiese sido el importe económico; debido a que se paga por volumen de almacenamiento y por volumen de texto con los que se utiliza la API *Natural Language*.

En lo que respecta al análisis de sentimiento de la palabra *Neuromarketing*, podemos concluir que su uso negativo es mínimo, y que en su mayoría se utiliza de forma neutra y positiva. Analizando con detalle los dos únicos tuits negativos, observamos que es debido a que en uno de ellos hace referencia a la pandemia y el otro a que el individuo no es consciente de lo que quiere comprar. Por lo que hay que tener en cuenta, que un texto que sea considerado negativo no significa que la palabra que estamos analizando tenga una connotación negativa; ya que en este caso en concreto, no se está hablando negativamente del neuromarketing sino de la no consciencia de los individuos en su toma de decisiones.

El hecho de que en todo el territorio nacional de España y que, durante el periodo de un mes, tan solo se detectan 209 usuarios que incluyen en su tuit la palabra *neuromarketing*, confirma que no es un sector en el que se genere mucho contenido en dicha red social. Hay que tener en cuenta, que gran parte del contenido es compartido debido al alto nivel de retuits existentes. Si medimos los tuits únicos, nos encontramos con que en un mes, tan solo se han generado 295 tuits, no llegando a una media de 10 tuits diarios.

Hay que destacar el área profesional de su uso, ya que su difusión se centra en su mayoría de lunes a viernes y en horario de 9:00 de la mañana a 17:00 de la tarde, aunque es cierto que no existe una relación directa entre los días y las horas de publicación, con respecto a la polaridad del contenido. El contenido tiende a ser mayoritariamente neutro, y es lógico debido al carácter informativo y de formación que tienen los tuits. Analizando los distintos *hashtags* que se han utilizado, podemos confirmar en primer lugar, que la presencia de la palabra *neuromarketing* no es transitoria. De lo contrario, no contaría con un porcentaje tan alto de *hashtag* que incluyen *#neuromarketing*. Por otro lado, se refuerza el uso profesional al estar acompañado de otras disciplinas del marketing; como pueden ser marketing digital y publicidad; y en menor medida pero también repetido en más de una ocasión, neurociencia, tendencias, copywriting, marketing emocional o neuroventas.

6. REFERENCIAS

- Agavanakis, K. N., Karpetas., G. E., Taylor. M., Pappa, E., Michail, C. M., Filos, J., Trachana, V., & Kontopoulou, L. (2019). Practical machine learning based on cloud computing resources. AIP Conference Proceedings, 2123. <https://doi.org/10.1063/1.5117023>
- Alcover de la Hera, C. M. (2008). Neurociencia social: hacia la integración de las explicaciones sociales y biológicas de la conducta social. Método, teoría e investigación en Psicología social. 187-214. Pearson Educación.
- Boubela, R. N., Kalcher, K., Huf, W., Našel, C., & Moser, E. (2016). Big Data Approaches for the Analysis of Large-Scale fMRI Data Using Apache Spark and GPU Processing: A Demonstration on Resting-State fMRI Data from the Human Connectome Project. *Frontiers in Neuroscience*, 9(JAN), 492. <https://doi.org/10.3389/fnins.2015.00492>
- Challita, S., Zalila, F., Gourdin, C., & Merle, P. (2018). A precise model for Google cloud platform. Proceedings - 2018 IEEE International Conference on Cloud Engineering, IC2E 2018, 177-183. <https://doi.org/10.1109/IC2E.2018.00041>
- Eklund, A., Andersson, M., & Knutsson, H. (2012). FMRI analysis on the GPU- Possibilities and challenges. *Computer Methods and Programs in Biomedicine*, 105(2), 145-161. <https://doi.org/10.1016/j.cmpb.2011.07.007>
- Harmon-Jones, E., & Beer Jennifer. (2009). *Methods in Social Neuroscience*. The Guilford Press.
- Liu, X., Hao, L., & Yang, W. (2019). Bigeo: A foundational PaaS framework for efficient storage, visualization, management, analysis, service, and migration of geospatial big data—a case study of Sichuan province, China. *ISPRS International Journal of Geo-Information*, 8(10). <https://doi.org/10.3390/ijgi8100449>

Quintana Gómez, A.

Análisis de los procesos de tratamiento de información en un estudio de análisis de sentimiento utilizando la tecnología de Google

Mart, E., Mart, M. T., y Ure, L. A. (2014). Desafíos del Análisis de Sentimientos. October. V Jornadas TIMM. 61–63.

Martínez Herrador, J. L., Núñez Cansado, M., y Valdunquillo Carlón, M. I. (2020). Metodología de neuromarketing: medición de Sociograph aplicada al análisis de la narrativa audiovisual erótica y sus aplicaciones a la estrategia de mercadotecnia. Vivat Academia. 131–153. <https://doi.org/10.15178/va.2020.150.131-153>

Poggi, N., Berral, J. L., Fenech, T., Carrera, D., Blakeley, J., Minhas, F., & Vujic, N. (2016). The state of SQL-on-Hadoop in the Cloud. 4th IEEE International Conference on Big Data (Big Data), 1432–1443. <https://doi.org/10.1109/BigData.2016.7840751>

Quintana, Á. (2019). Contrastando herramientas del Proceso del Lenguaje Natural con encuestas. La nueva comunicación del siglo XXI. Pirámide

Singh, J., Gill, R., & Goyal, G. (2019). Extracting and understanding user sentiments for big data analytics in big business brands. In Big Data Recommender Systems - Volume 2: Application Paradigms. Institution of Engineering and Technology. 235-257. Institution of Engineering and Technology.

Yuhanna, N. (2020). The Forrester Wave™: Data Management For Analytics, Q1 2020 The 14 Providers That Matter Most And How They Stack Up Key Takeaways. The Forrester <https://reprints.forrester.com/#/assets/2/157/RES157286/reports>

AUTOR

Ángel Quintana Gómez

Licenciado en Ciencias de la Información, rama de Publicidad, Relaciones Públicas por la Universidad Complutense de Madrid. Doctor en Comunicación Audiovisual, Publicidad y Relaciones Públicas. En la actualidad, es profesor en el Grado de Comunicación en la Universidad del Atlántico Medio en Gran Canaria, donde también imparte asignaturas en el Título Superior de Marketing y Negocios Digitales. Director del Máster en Comunicación, Publicidad y Marketing de Gran Canaria y Tenerife que imparte dicha Universidad.

Orcid ID: <https://orcid.org/0000-0002-5139-8755>

Google Scholar:

<https://scholar.google.es/citations?user=AoERoAIAAAAJ&hl=es&oi=ao>