



Facultad de  
**Comunicación y Documentación**

UNIVERSIDAD DE GRANADA

GRADO EN INFORMACIÓN Y DOCUMENTACIÓN

TRABAJO FIN DE GRADO

**INTRODUCCIÓN A LAS TÉCNICAS DE ANÁLISIS DE  
SENTIMIENTOS**

Presentado por:

**D./D<sup>a</sup>. Sara Huergo Gonzalvo**

Tutor:

**Prof. Dr./Dra. o D./D<sup>a</sup> Enrique Herrera Viedma**

Curso académico 2019 / 2020



D./Dña.: Enrique Herrera Viedma, tutor/a del trabajo titulado **Introducción a las Técnicas de Análisis de Sentimientos** realizado por el alumno/a **Sara Huergo Gonzalvo**, INFORMA que dicho trabajo cumple con los requisitos exigidos por el Reglamento sobre Trabajos Fin del Grado en *Información y Documentación* para su defensa.

Granada, 18 de noviembre de 2019

A handwritten signature in black ink, appearing to be 'EHV', written in a cursive style.

Fdo.: Enrique Herrera Viedma



Por la presente dejo constancia de ser el/la autor/a del trabajo titulado **Introducción a las Técnicas de Análisis de Sentimientos** que presento para la materia Trabajo Fin de Grado del Grado en **Información y Documentación**, tutorizado por el/la profesor/a **Enrique Herrera Viedma** durante el curso académico 2019 - 2020.

Asumo la originalidad del trabajo y declaro que no he utilizado fuentes (tablas, textos, imágenes, medios audiovisuales, datos y software) sin citar debidamente, quedando la Facultad de Comunicación y Documentación de la Universidad de Granada exenta de toda obligación al respecto.

Autorizo a la Facultad de Comunicación y Documentación a utilizar este material para ser consultado con fines docentes dado que constituyen ejercicios académicos de uso interno.

**18 / 11 / 2019**

Fecha



Firmado: Sara Huergo Gonzalvo



# ÍNDICE

RESUMEN.....	10
Abstract .....	10
1. INTRODUCCIÓN .....	12
2. OBJETIVOS.....	13
3. METODOLOGÍA .....	13
4. EL ANÁLISIS DE SENTIMIENTOS .....	14
4.1. PROCESAMIENTO DEL LENGUAJE NATURAL (PLN).....	14
4.2. TAREAS DEL ANÁLISIS DE SENTIMIENTOS.....	16
4.3. TÉCNICAS DE ANÁLISIS DE SENTIMIENTOS.....	18
4.3.1. ESTRATEGIAS BASADAS EN APRENDIZAJE AUTOMÁTICO ( <i>MACHINE LEARNING</i> )	18
4.3.2. ESTRATEGIAS BASADAS EN DICCIONARIOS ( <i>LEXICON-BASED</i> ).....	19
4.4. PROBLEMAS Y LIMITACIONES DEL ANALISIS DE SENTIMIENTOS.....	20
5. HERRAMIENTAS.....	21
5.1. <i>SOCIALMENTION</i> Y <i>WHO'S TALKING</i> .....	21
5.2. <i>RAPIDMINER</i> .....	23
5.3. <i>OPINION CRAWL</i> .....	23
5.4. <i>CHATTERSCOPE</i> .....	24
5.5. <i>QUICK SEARCH</i> .....	24
5.6. <i>HOOTSUITE INSIGHTS</i> .....	25
5.7. <i>BRAND24</i> .....	26
6. EXPERIMENTOS .....	27
6.1. <i>RAPIDMINER</i> .....	27
6.1.1. OPERADORES NECESARIOS.....	28
6.1.2. RESULTADOS .....	32
6.2. <i>BRAND24</i> .....	33
6.2.1. <i>MENTIONS</i> : escritorio principal del análisis.....	34
6.2.2. <i>SUMMARY</i> .....	37

6.2.3.	<i>ANALYSIS</i> .....	37
6.2.4.	<i>COMPARISON</i> .....	37
6.2.5.	<i>PROJECT SETTINGS</i> .....	38
7.	CONCLUSIONES .....	39
	BIBLIOGRAFÍA.....	41





## **RESUMEN**

El análisis de sentimientos es la técnica capaz de detectar los sentimientos subjetivos que se encuentran tras los comentarios que los usuarios publican en las redes sociales, foros, blogs, etcétera, y clasificar dichos sentimientos detectados según sean positivos, negativos o neutros. Todo esto se consigue a través de una serie de modelos y técnicas que enseñan al ordenador a entender el lenguaje natural utilizado por los usuarios en la web e identificar los sentimientos con los que se han escrito las publicaciones.

En este trabajo se exponen estos conceptos y algunos más que son necesarios para entender en profundidad cuál es el funcionamiento del análisis de sentimientos y qué limitaciones son las que impiden que su eficacia sea total. Además, se exponen algunas herramientas que existen (o existieron) para realizar tal tarea y se mostrará la utilización de dos de estas herramientas para comprobar su funcionamiento con la red social Twitter.

### **Abstract**

Sentiment analysis is one technique made for detect the subjective feelings that can be find at the comments published by the users in social networks, forums, blogs, etc., and classify these sentiments according to whether they are positive, negative or neutral. All of this is achieved through a series of models and techniques that can teach the computer how to understand the natural language used by the users on the web and identify the feelings with which the publishes were wrote.

At this work, some concepts like this are exposed and some more necessary to understand what is the functioning of the sentiment analysis and what are those limitations that prevents its effectiveness from the total. Even more, some tools that exists (or existed) to make this work are exposed and two of those tools will be shown to prove how it works with the social network Twitter.



# 1. INTRODUCCIÓN

El crecimiento de la Web 2.0 y el uso masivo de las redes sociales son las razones principales de la puesta en marcha de tareas como el análisis de sentimientos. A día de hoy, la mayoría de las empresas, compañías, organizaciones, etc., tienen presencia en la web y es aquí donde sus consumidores producen una gran cantidad de comentarios y exponen públicamente su opinión y sentimientos sobre los productos, empresas, y entidades o incluso sobre temas más generales como la cultura, la política, la sociología o la economía, entre otros. Estos textos resultan muy valiosos para dichas empresas o entidades pues es esa percepción pública la que los puede ayudar a mejorar, a tomar decisiones o a conocer cuáles son las necesidades de sus usuarios o incluso a identificar amenazas u ofensas. Como ejemplo ilustrativo se puede nombrar el uso de esta técnica en Twitter para la predicción de los resultados en unas elecciones políticas o para conocer el recibimiento y aceptación que tienen los consumidores ante ciertos productos o ante una marca.

Esta información a la que llamamos “útil”, no solo se encuentra en el texto en sí escrito por el usuario, sino también en los sentimientos y emociones subjetivos que hay implícitos en el mismo. Esta es la razón de que el análisis de las emociones y sentimientos que hay detrás de los comentarios sea mucho más importante incluso que analizar los textos por su estructura lingüística. De tal forma, el análisis de sentimientos combina técnicas de minería de texto u opinión (opinion mining) y Procesamiento del Lenguaje Natural (PLN) para acceder a la información deseada, aunque presenta ciertas dificultades y problemas que hacen que la precisión con la que aporta los resultados del análisis no sea del 100% pues la simple lectura del lenguaje natural que utilizan los usuarios en las redes sociales, blogs, foros, etcétera, para expresar su opinión, es fácil de entender por otro usuario pero difícil para el lenguaje de una máquina como es un ordenador.

En este trabajo se esclarecen dichos problemas y técnicas del AS. El proyecto se clasifica en tres grandes bloques sin contar esta introducción (que da una presentación general al tema tratado), los objetivos (donde se expone cuál es la finalidad de este proyecto y qué se quiere aportar con él) o la metodología (que responderá a cómo se ha realizado el trabajo). En el primer bloque, se tratan los conceptos básicos para entender el funcionamiento y definición del análisis de sentimientos y las limitaciones y problemas que presenta además de las estrategias más comunes utilizadas en su funcionamiento. En una segunda sección, se hace

una revisión de algunas de las herramientas que existen actualmente para realizar este tipo de tarea y seguido, en una tercera sección, una serie de experimentos utilizando dos de los programas expuestos en la revisión anterior. En esta sección práctica se explica, a través de texto y fotografías que ilustren la explicación y hagan más sencilla su comprensión, el funcionamiento de dichas herramientas para realizar un análisis de sentimientos lo más completo posible. Como último punto del trabajo se expondrán las conclusiones generales del trabajo y se incluirá la bibliografía utilizada.

## **2. OBJETIVOS**

El objetivo general de este trabajo es hacer una introducción al análisis de sentimientos, exponiendo los conceptos básicos necesarios para entender qué es esta técnica, que funciones tiene, para qué se utiliza y cómo se utiliza, para lo que se incluyen algunos ejemplos prácticos que visualizaran la respuesta a esta cuestión.

## **3. METODOLOGÍA**

Para las diferentes partes de este trabajo se han utilizado metodologías diferentes. Para la parte teórica se ha realizado una búsqueda exhaustiva de la bibliográfica especializada en este tema con la finalidad de extraer la información básica para la comprensión del análisis de sentimientos. Para la segunda sección se ha realizado la lectura de algunos artículos que incluyen revisión de herramientas para el análisis de sentimientos y se ha hecho la búsqueda en internet de dichas herramientas y el estudio de sus funcionalidades desde las páginas web principales de las mismas. Para la última sección, parte práctica de este trabajo, se han utilizado dos herramientas (RapidMiner y Brand24) para realizar un análisis de sentimientos básico y conocer así la funcionalidad que se pueden encontrar en este tipo de herramientas.

## **4. EL ANÁLISIS DE SENTIMIENTOS**

El análisis de sentimientos es a la vez una tarea y un problema del PLN. Como tarea se encarga de identificar las opiniones y actitudes de los usuarios relacionadas con entidades, individuos, eventos, etc., es decir, identificar los sentimientos que se esconden tras los textos y comentarios escritos en la web. Como problema es limitado pues todas sus acciones, sumadas a las de la Recuperación de Información (RI), luchan continuamente contra los problemas sin resolver del PLN, esenciales para detectar recursos literarios como la ironía o el sarcasmo y, por ende, encontrar y calificar sentimientos (Liu, 2006).

Tripathi y S. Naganna (2015) precisaron más en su definición del análisis de sentimientos exponiendo que es el proceso de extraer, identificar, analizar y caracterizar los sentimientos u opiniones de la información textual usando aprendizaje automático, PLN o métodos estadísticos. Marcaban así los tres pasos clave a realizar en un análisis de sentimientos básico: primero, identificar el sentimiento expresado en el texto; segundo, identificar al propietario del sentimiento y la entidad a la que va dirigido; tercero, identificar la polaridad del sentimiento.

Estas son tan solo dos de las muchas de definiciones que se hacen del concepto de AS debido a que es una técnica en continua investigación para su actualización y resolución de los problemas y limitaciones que tiene. Pero antes de conocer cuáles son estos problemas sin resolver del análisis de sentimientos o cuales son las tareas que realiza, se debe definir bien qué es el PLN y cuál es su función pues es uno de los puntos más importantes del AS.

### **4.1. PROCESAMIENTO DEL LENGUAJE NATURAL (PLN)**

Esta técnica se utiliza para conseguir que los ordenadores puedan entender el lenguaje convencional usado por los usuarios, procesarlo y realizar así cálculos y análisis descriptivos o análisis de sentimientos. El uso de este proceso se debe a la dificultad de la tarea de análisis debido, en primer lugar, al volumen ingente de comentarios que se publican diariamente en la web, que hace imposible el análisis de forma manual o física, y, en segundo lugar, el registro informal que utilizan los usuarios en sus comentarios. Así pues, a través del análisis de texto, el PLN “limpia” los comentarios generados por los usuarios y los transforma al formato necesario para que puedan ser entendidos por el ordenador y así poder extraer la información deseada. Este proceso incluye una serie de pasos:

- Una limpieza general del texto original que consiste en eliminar los signos de puntuación y convertirlo todo a minúsculas.
- Tras esto se da paso al *Stop Words Removal*, un borrado de “palabras vacías”, es decir, las palabras más frecuentes que no aportan información sobre el sentimiento del tweet (pronombres, artículos...).
- El *Stemming* es el paso que se encargará de acortar las palabras restantes a su raíz para simplificar los datos de los mensajes. Medhat (2014) también habla de lematizar, es decir, identificar palabras con el mismo significado semántico y unificarlas. Por ejemplo, las palabras en plural convertirlas al singular.
- Por último, se realiza una lista de palabras a partir del mensaje y se tratan todas de la misma forma.

Por ejemplo:

- **Tweet Original:** “@apple Quiero vender MI IPHONE 6 PLUS!!!! y comprarme el iphone 7... Lo quieren XD?”
- **Limpieza:** @apple quiero vender mi iphone 6 plus y comprarme el iphone 7 lo quieren
- **Stop Words Removal:** @apple quiero vender iphone 6 plus comprarme iphone 7 quieren
- **Stemming:** @apple querer vender iphone 6 plus comprar iphone 7 querer
- **Lista de palabras:** [@apple] [querer] [vender] [iphone] [6] [plus] [comprar] [iphone] [7] [querer]

Una vez realizado este pre-procesamiento de los datos, se pasa al análisis de sentimientos en sí entrenando un clasificador para clasificar los mensajes según su polaridad, es decir, si son positivos o negativos o neutros. (BD Guidance, 2016)

Uno de los aspectos principales con los que tiene que lidiar el PLN, son los diferentes niveles de análisis: el documento, la oración y la entidad.

- El análisis a nivel de **documento** considera la totalidad de este como una unidad que ofrece información sobre un tema y expresa opiniones y/o sentimientos positivos o negativos sobre el mismo.
- El análisis a nivel de **oración** identifica primero si las oraciones del texto son objetivas o subjetivas y en este último caso, detecta si la opinión es positiva o negativa. Theresa

Wilson (2005) no diferencia entre este nivel y el anterior pues considera las frases como documentos cortos que no aportan información detallada sobre una opinión o sentimiento. Por ello utiliza el nivel de entidad.

- El análisis a nivel de **aspecto** clasifica los sentimientos respecto a características o rasgos específicos de un producto o entidad. Es decir, primero identifica la entidad sobre la que se ofrece la opinión y después sus aspectos más característicos sobre los que los usuarios exponen sus opiniones y sentimientos. Este nivel es el más complejo y por tanto requiere métodos más complejos. (Medhat, 2014)

Una vez entendido este proceso será más fácil entender como trabaja el análisis de sentimientos.

## 4.2. TAREAS DEL ANÁLISIS DE SENTIMIENTOS

Para entender correctamente el funcionamiento de las tareas de este análisis, se debe diferenciar con claridad entre la opinión y el sentimiento o emoción y entre la información objetiva y subjetiva. Jesús Serrano et al. (2015), en su artículo “*Sentiment Analysis: A review and comparative analysis of web service*”, definen muy claramente estos conceptos.

- **Detección de opiniones y emociones/sentimientos:** La opinión es el concepto transitorio que refleja una actitud evaluable (positiva o negativa) sobre un producto, evento, organización, etc. Pueden ser directas, si son ideas sobre una entidad o aspecto de la misma, o indirectas, si la opinión sobre la entidad está basada en los efectos de otras entidades u opiniones.

Las **emociones**, por el contrario, son la expresión de los **sentimientos** subjetivos. Walaa Medhat (2014), en su artículo *Sentiment Analysis algorithms and applications: A survey*, también habla de la detección de emociones (*Emotion Detection*, ED) y su profundidad en el análisis de sentimientos exponiendo que esta última es una tarea del PLN y solo puede detectar opiniones y evaluarlas entre positivo o negativo mientras que la ED es una tarea dentro del AS capaz de detectar las emociones que se esconden detrás de un texto (alegría, tristeza, ira, miedo, sorpresa...).

- **Clasificación de subjetividad y objetividad:** consiste, como su propio nombre indica, en detectar y clasificar si un texto es subjetivo u objetivo. La diferencia entre ambos conceptos es clara: mientras que la objetividad se basa en hechos reales iguales a cualquier visión, la subjetividad implica un sentimiento positivo o negativo pues es



la expresión de puntos de vista absolutamente personales. Es decir, es la visión personal del usuario contra la información fehaciente que se tiene sobre el tema o entidad. La clasificación subjetiva es más difícil de realizar que la objetiva, pero asegura una mejor clasificación de los sentimientos.

- **Recuperación y resumen de opiniones:** son tareas previas a la detección de sentimientos en sí. La recuperación se encarga de encontrar aquellos documentos que expresan una opinión sobre una consulta dada y en la fase de **resumen** se extraen las características principales de una entidad según lo incluido en uno o varios documentos. Si el resumen se hace de un solo documento, se estudian los cambios de sentimientos que ocurren a lo largo del mismo. Sin embargo, si el resumen es de varios documentos, lo que se hace es una clasificación de las frases (positivas o negativas) que expresan sentimientos relacionados con las entidades, creando un gráfico final que muestre los rasgos característicos de las entidades y cuantifique los sentimientos.
- **Clasificación y calificación de sentimientos:** es la tarea principal del análisis de sentimientos, clasificar y calificar los sentimientos detectados en los comentarios según sus intenciones o intensidad de la emoción. Las variables encargadas de medir esto son (Bravo-Marquez et al., 2014):
  - **Polaridad:** clasifica los mensajes en positivos, negativos o neutros. Esta es la variable más sencilla de entender y con la que más se trabaja en este tipo de análisis.
  - **Intensidad:** proporciona un valor numérico dependiendo de la intensidad del sentimiento.
  - **Emoción:** esta tercera variable, menos utilizada, clasifica el texto según los distintos tipos de emociones.

Las técnicas de análisis de sentimiento son las que se encargan de realizar toda esta clasificación y se pueden dividir en dos grupos: son las técnicas basadas en aprendizaje automático (*machine learning approach*) y las que se basan en diccionarios (*lexicon-based approach*).

### 4.3. TÉCNICAS DE ANÁLISIS DE SENTIMIENTOS

#### 4.3.1. ESTRATEGIAS BASADAS EN APRENDIZAJE AUTOMÁTICO (*MACHINE LEARNING*)

Se trata de un programa informático que se encarga de que el ordenador extraiga unos patrones de comportamiento a partir de entradas recibidas y de la selección y extracción de una colección de rasgos usados para detectar sentimientos y, en base a esa información aprendida, realice la evaluación de nuevas entradas. Consiste en el entrenamiento de un clasificador usando algoritmos internos, con componentes estadísticos y algebraicos. (Baviera, 2017). Aquí el PLN desempeña una función muy importante pues los rasgos que más se tienen en cuenta suelen ser: los términos y la frecuencia de su uso, la categoría gramatical de los términos (adjetivos, sustantivos), negaciones que cambian el significado de la frase y las dependencias sintácticas. (Arcila-Calderón, 2017)

Esta técnica se divide en dos estrategias dependiendo de la retroalimentación: supervisadas y sin supervisión. Las primeras son las más usadas y utilizan una gran cantidad de documentos de entrenamiento creados y clasificados semiautomáticamente para realizar predicciones del tipo de sentimiento expresado en nuevos textos a clasificar. Mientras, las técnicas de aprendizaje automático no supervisado se utilizan cuando es difícil encontrar dichos documentos clasificados. Por ello utiliza algoritmos de agrupamiento o “*clustering*” que se encargan de buscar patrones de semejanza en los textos y asociarlos por sus características comunes. (Medhat, 2014)

El punto más importante y primer paso del aprendizaje automático supervisado, es el entrenamiento de la máquina a partir de la elaboración de dicho corpus de documentos clasificados. Dicho corpus tiene una doble función y por ello se divide en dos grupos: por un lado, el conjunto más grande que corresponde al corpus de desarrollo y que se encarga del aprendizaje y ajuste del algoritmo y por otro, el corpus de prueba que se encarga de evaluar su fiabilidad y calcular la precisión del algoritmo ya entrenado. La segunda fase es la de clasificación que aplica el algoritmo ya ajustado y entrenado para procesar el gran volumen de datos. Es en este proceso cuando se realiza la clasificación de los mensajes según su polaridad. Eso sí, este entrenamiento y clasificación dependen del contexto en el que se aplican. Es decir, un algoritmo entrenado para clasificar textos de un ámbito político no se podría aplicar para clasificar la polaridad de un ámbito deportivo. (Baviera, 2017)

El elemento común a ambas fases es el extractor de atributos, un módulo que se encarga de evaluar las características del texto. Dependiendo de la fase en la que se encuentre el proceso, el algoritmo usará estas características de forma distinta: si se encuentra en la fase de entrenamiento servirán para eso mismo, entrenar al algoritmo, y si se encuentra en la fase de clasificación se utilizará para categorizar el documento. (Baviera, 2017)

Dentro de este módulo se encuentra el preprocesamiento de datos que se encarga de “limpiar y ordenar”, proceso perteneciente al PLN que ya se ha explicado anteriormente en el apartado dedicado a este proceso.

#### **4.3.2. ESTRATEGIAS BASADAS EN DICCIONARIOS (*LEXICON-BASED*)**

Esta es una alternativa basada en diccionarios y algoritmos más sencillos. Depende de un conjunto de términos asociados a los sentimientos e incluso frases hechas o expresiones desarrolladas para géneros de comunicación tradicionales o incluso algo más complejos como las ontologías. Podemos clasificar esta estrategia en dos enfoques, uno semántico caracterizado por el uso de diccionarios de sentimientos, también conocidos como *lexicons*, y otro basado en el corpus del documento. (Medhat, 2014)

El enfoque basado en un diccionario surge con el objetivo de crear colecciones de términos relacionados con opiniones y sentimientos. Funciona usando un conjunto inicial de términos de opinión que se recopilan de forma manual y crece a través de la búsqueda de sinónimos y antónimos y palabras relacionadas. Esta estrategia permite que el ordenador procese, reconozca y evalúe las emociones contenidas en el mensaje. Dependiendo del idioma en el que se escriban los diccionarios, podemos encontrar diferentes orientaciones para clasificar los sentimientos. Por ejemplo, los escritos en español, están orientados a la polarización de los términos como son Elhuyar y el CRiSOL.

El corpus comienza con una lista de palabras de opinión y después encuentra otras en el corpus del documento para ayudar a identificar la opinión en sí con un contexto específico. Esto se hace usando métodos estadísticos (para establecer la polaridad de los sentimientos) o semánticos. (Medhat, 2014)

El algoritmo de este proceso tan solo detecta coincidencias con los diccionarios y evalúa el sentimiento según las coincidencias encontradas. Esta es una tarea mucho más sencilla que en

el que caso anterior pues aquí, el algoritmo, no tiene que detectar patrones sintácticos ni aprender para después realizar la tarea.

#### **4.4.PROBLEMAS Y LIMITACIONES DEL ANALISIS DE SENTIMIENTOS**

Anteriormente, en el apartado de explicación del PLN, se han expuesto las ya limitaciones claras del AS, como es, por ejemplo, el uso del lenguaje natural por los usuarios en las redes sociales y que el ordenador o máquina no puede entender. Esto hace que la precisión del análisis de sentimientos no pueda ser del 100%. Según Li (2015), en 2015 la precisión llegaba a ser del 70%, un muy buen resultado que sin embargo hay que seguir incrementando hasta conseguir el máximo posible.

Algunos de los problemas y dificultades que presenta el análisis de sentimientos son:

- a) **Diferentes niveles de sentimientos.** El sentimiento de una palabra no depende simplemente del significado en sí de la misma, sino que puede cambiar por el contexto en el que se encuentra. Además, existen muchas palabras que implican el mismo sentimiento, pero con diferentes grados de intensidad. Por ejemplo, “no está mal”, “bien” y “muy bien”, son tres expresiones que expresan un sentimiento positivo, pero con diferente intensidad. Sin embargo, la primera de ellas podría clasificarse como negativa si el programa solo reconoce la palabra “mal”. (Liu, 2006)
- b) **Contexto de los adjetivos.** De nuevo, la importancia de reconocer el contexto en el que se encuentran las palabras para identificar el sentimiento. Las palabras más susceptibles de producir un error son los adjetivos, cuya polaridad cambia dependiendo del contexto en el que se encuentren. (Chiara, 2018)
- c) **Diferenciar la objetividad y subjetividad.** Tal y como se ha explicado en el apartado del PLN, uno de los límites del AS es la diferenciación entre la opinión objetiva contrastada con hechos reales o la opinión subjetiva donde entran los sentimientos de cada una de las personas por individual, lo que influye en la polaridad del comentario que publica y que al ordenador le cuesta identificar.
- d) **Dificultad para detectar las opiniones implícitas o explícitas.** La opinión explícita es una opinión simple y comparativa que no expresa más allá de lo que se encuentra en la información textual. Al contrario, la opinión implícita es aquella en la que el autor del texto no expone de forma directa su opinión de tal forma que habrá que leer

entre líneas para captar la información, algo que un ser humano puede hacer, pero no una máquina. Este caso va muy unido a la dificultad para entender el sarcasmo o la ironía.

## **5. HERRAMIENTAS**

Ya se ha comentado lo importante que es para las empresas y marcas monitorizar lo que los usuarios dicen sobre ellas en la red. Ya no solo las menciones o cuándo se comenta sobre ellas sino, cómo se comenta, cuál es el contenido del mensaje y valorar si es positivo, negativo o neutro. En resumen, la monitorización de los sentimientos. Para ello existen una gran cantidad de herramientas, pero, el continuo estudio y avance de este tipo de proceso hace que muchas de ellas, se queden obsoletas al poco tiempo y sean intercambiadas por otras que incluyen más funcionalidades o inmediatez en el análisis. De hecho, la mayoría de las herramientas de las que podemos leer artículos y revisiones en la web, surgieron en el año 2010 y a día de hoy, muchas, ya no existen. Otras, sin embargo, están actualizadas e incluyen muchísimas más funciones muy útiles para realizar un análisis completo.

Un ejemplo, una vez más, es el artículo de Jesús Serrano (2015), ya citado anteriormente, en el que hace una revisión de algunas de dichas herramientas, pero muchas ya no existen o no funcionan. Llama la atención como, en tan solo cuatro años, muchas de las herramientas han quedado inservibles y reemplazadas por otras nuevas que actualmente tienen muchas más funcionalidades que las de entonces. Sin embargo, algunas de las que expusieron en el artículo citado, siguen en funcionamiento y con acceso semi-abierto, permitiendo una descarga e instalación de prueba durante 30 días para después pasar a la versión ilimitada de pago.

En este apartado del trabajo se verá una relación de algunas de estas herramientas y, en el apartado siguiente, parte práctica del trabajo, se verá la utilización de algunas de ellas como demostración.

### **5.1. SOCIALMENTION Y WHO'S TALKING**

Social Mention (<http://socialmention.com/>) es un buscador web que funciona tal y como cualquier otro buscador: a través de consultas realizadas con términos o palabras clave que actuarán de filtro para mostrar los lugares en los que aparece dicha consulta además de indicadores estadísticos que ayudaran a la empresa o entidad a conocer la imagen que se tiene

de ella en la red. Es decir, su misión final es analizar todos los contenidos agregados por los usuarios en internet sobre la marca o empresa, monitoreando las menciones (quién), el lugar (dónde) y el momento de publicación (cuándo). (González, 2011)

Para medir la influencia que tienen las marcas, eventos o asuntos en los blogs, redes sociales, etcétera, utiliza unos parámetros que se encargan de determinar estos valores: la fuerza, los sentimientos (positivos o negativos como se ha dicho anteriormente), la pasión (la participación de la misma persona más de una sola vez) o el alcance (el rango de influencia de los autores). (Velasco, 2010)

Entre los servicios que ofrece destacan la API para desarrolladores o la descarga de un archivo Excel como almacenamiento de menciones y resultados. Pero como cualquier otra herramienta, tiene ciertas limitaciones. La más destacable es la asignación de sentimientos, la cual hace de forma automática y bajo unos criterios que no se pueden cambiar, lo que significa que el usuario no puede configurar lo que considera positivo o negativo ni podrá corregir los errores típicos de la automatización.

Su punto fuerte es la monitorización a tiempo real y la interacción con las plataformas desde un solo tablero directo. Pero, además, ofrece un tipo de monitorización diferente y particular para cada una de las plataformas. Esto es bastante visible en el trabajo que realiza con las tres redes sociales más importantes: Twitter, Facebook e Instagram.

En Twitter, por ejemplo, analiza las conversaciones profundizando en el significado y puntos de vista diferentes además de clasificar comentarios según los sentimientos o eliminar el ruido. En Facebook, que es una plataforma donde los posts se pueden perder fácilmente en el tablón principal, reacciona inmediatamente a las novedades y conseguirá que nunca se pierda un post que sea interesante para el cliente. Para Instagram, añade una monitorización de *hashtag* y visualizaciones de las menciones.

***Who's Talkin'*** (<https://www.whostalkinseo.com/>) es también un buscador que muestra los sitios donde se encuentra el término buscado, tal y como la herramienta anterior, razón por la que se las ha colocado juntas en esta revisión de herramientas. Su diferencia característica con *SocialMention*, es que ofrece la posibilidad de exportar una hoja de cálculo con los resultados obtenidos de la búsqueda. (Velasco, 2010)

## 5.2. RAPIDMINER

RapidMiner (<https://rapidminer.com/>) es un software de minería de datos que incluye técnicas de procesamiento de datos, modelación predictiva, visualización de datos, etcétera. Es la segunda herramienta más usada por las grandes compañías de cerca de 40 países según la encuesta anual realizada en el portal internacional de Minería de Datos KDnuggets (<https://www.kdnuggets.com/>).

Tal y como el mismo equipo del software define en su página web, RapidMiner unifica el ciclo completo de vida de la ciencia de datos, desde la preparación y pre-procesamiento de datos para el aprendizaje automatizado hasta la utilización del modelo predictivo.

Algo muy interesante de este software y su equipo es que, en su página web oficial, ofrecen una plataforma específica para la formación y entrenamiento con la herramienta en la que incluyen una serie de tutoriales en vídeo y texto y explicaciones para que el uso de la misma sea fácil de entender y realizar. Todo esto se puede encontrar desplegando la pestaña “Learn” y entrando en *RapidMiner Academy*. Pero esto es solo una pequeña parte de todo lo que puede ofrecer dicha herramienta.

En la parte práctica de este trabajo se explicará de manera sencilla y clara el uso y el proceso a seguir para programar el software haciendo un análisis de sentimientos de la red social Twitter.

## 5.3. OPINION CRAWL

Es una herramienta creada por Semantic Engines LLC, una compañía privada que utiliza técnicas de minería de texto, análisis semántico o análisis de sentimientos. Una de sus herramientas más importantes es *SenseBot*, un buscador semántico online que genera un resumen de los resultados recuperados sobre una búsqueda realizada por el usuario.

Opinion Crawl (<http://www.opinioncrawl.com/>) es otra plataforma de esta compañía que, al igual que la anterior, consiste en un buscador online pero que en este caso evalúa los sentimientos de diferentes temas extraídos de las redes sociales como Twitter. Cuando se realiza una búsqueda, la información que se obtiene

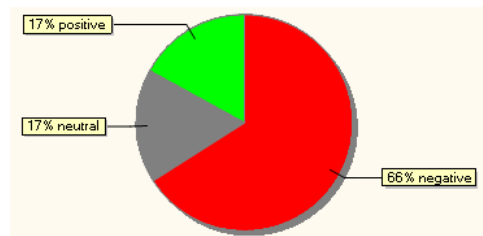


Fig. 1. Gráfica generada con Opinion Crawl

incluye: una gráfica mostrando los sentimientos actuales sobre el tema en tiempo real (Fig.1),

una lista de los últimos titulares publicados, algunas imágenes sobre el tema y una nube de palabras clave que los usuarios suelen relacionar con la consulta realizada. Todo esto en su conjunto permite comprobar cuáles son los temas principales que pueden hacer que la gráfica se vea de una forma u otra teniendo en cuenta el análisis de sentimientos.

#### **5.4. CHATTERSCOPE**

Chatterscope es una plataforma gratuita desarrollada por Lewis PR. en 2010. Permite la monitorización y análisis de sentimientos en Twitter, registrando las menciones que se hacen de las marcas, empresas, entidades, etc., y clasificando automáticamente los tweets haciendo un porcentaje de positivos, negativos o neutros. Estos porcentajes se representan con gráficas visuales que muestran la evolución de los sentimientos a lo largo del tiempo. Todo esto en su conjunto ofrece una visualización del progreso de la reputación de la marca o entidad en tiempo real. (Velasco, 2010)

Algo muy interesante es que da la posibilidad de añadir palabras clave que se utilizarán para evaluar los mensajes/tweets. Además, ofrece la programación de alertas de las menciones que se hacen de la marca y que se podrán recibir cada hora, día o semana.

#### **5.5. QUICK SEARCH**

Es un buscador de redes sociales impulsado por *talkwalker* que ofrece una gran cobertura en redes, noticias, blogs y foros. *Quick Search* (<https://www.talkwalker.com/es/quick-search>) realiza un resumen instantáneo de la marca/entidad e impulsa el impacto de la misma en la web. (Carpenter, 2018)

Su interfaz ofrece información general sobre la marca como son la interacción, el volumen, los sentimientos, la demografía o la geografía. Pero además presenta elementos visuales para la monitorización de la marca en las redes sociales como son los gráficos de sentimientos (Fig.2). En ellos se muestra una barra de rango donde se observa por color el porcentaje de comentarios positivos y negativos.



Fig. 2. Diagrama de rangos comparativos de Quick Search

Otras funciones que ofrece y que son muy interesantes y útiles para analizar e impulsar la marca en las redes son:



- **Inteligencia competitiva:** analiza las millones de conversaciones que existen de la marca y contrasta la actividad e información de otras y hace una gráfica comparativa de ellas para obtener información rápida. Al ser un buscador de redes sociales, además, permite profundizar más y encontrar la fuente que produce sentimientos negativos para poder actuar rápida y eficazmente.
- **Ideación de contenido:** crea contenido viral que llegue al público a través de mensajes atractivos y contenido de tendencia y da un impulso al alcance de los mensajes propios mediante el uso de *emojis* o *hashtags*.

El uso de esta herramienta es el más sencillo. No se necesita ningún tipo de configuración ni tutoriales de uso. Aun así, ofrece una zona de formación en su plataforma, aunque está más destinada a dar unos conocimientos básicos sobre marketing. Una característica muy clara de esta herramienta es que su página web y plataforma están en varios idiomas, algo que no es común en este tipo de herramientas (la inmensa mayoría se encuentra en inglés) y que es muy útil y llamativo para las empresas interesadas en este tipo de trabajos. Pero presenta un inconveniente y es que no ofrece una prueba gratis durante un periodo de tiempo limitado, algo muy necesario y útil para que las empresas conozcan y tomen contacto con la herramienta y así se decanten por ella en vez de buscar alguna otra que cumpla con sus necesidades.

## **5.6. HOOTSUITE INSIGHTS**

*Hootsuite* (<https://hootsuite.com/products/insights#>) analiza automáticamente las plataformas de redes sociales, blogs, foros, etc. y revela la percepción y sentimientos de los usuarios ante otros usuarios, historias, marcas o entidades.

Permite filtrar por ubicación, idioma (más de 50) o género e incluye una monitorización a través de palabras clave que le ayudan a entender que sienten los usuarios y clasificar esto en positivo o negativo. Además, es capaz de monitorizar conversación en tiempo real y alertar de picos inusuales de volumen, sentimiento... (Carpenter, 2018)

Como la mayoría de herramientas, ofrece resultados visuales como son gráficas o dibujos para mayor comprensión de lo que ocurre en la web alrededor de la marca o entidad.

*Hootsuite*, al igual que la herramienta anterior, se encuentra en varios idiomas lo que hace que sea fácilmente accesible a cualquiera que requiera su idioma natal para entender a la

perfección este tipo de herramienta. Ofrece también una demostración de 30 días, pero con la diferencia de que, para acceder a ella, se debe incluir la cuenta bancaria donde se cargará el precio del programa al pasar los días de prueba.

## **5.7. BRAND24**

Es un *software* de supervisión de sitios web como son redes sociales, blogs, foros... Su función principal es identificar y analizar conversaciones en línea sobre las marcas, productos o entidades deseadas. Permite funciones de alerta para mantener a los usuarios actualizados sobre todo lo que ocurre en la red. (MarTech Forum, 2019)

Es un programa de pago pero como la inmensa mayoría, tiene una prueba gratuita por 14 días que permitirá utilizar la mayor parte (si no todas) de sus funciones, lo que permite a las grandes empresas decidirse a escogerla como su herramienta de análisis de datos. Además, no necesita ningún tipo de descarga en el ordenador pues es una versión que se encuentra en la nube.

Ofrece también formación con cursos en línea y documentación muy variada que permite que el uso de la herramienta sea muy sencillo. Aun así, tiene una interfaz bastante sencilla e intuitiva en la que es muy difícil perderse o fallar. Todos los resultados que obtiene y/o muestra están perfectamente titulados o colocados en diferentes lugares de tal forma que no den lugar a confusiones, algo muy útil sobre todo para los usuarios inexpertos.

La función de análisis de sentimientos es muy completa en este *software* pues clasifica las publicaciones según la polaridad pero permite cambiarla si leyendo se comprueba que ha habido una falta de comprensión por parte de la máquina. Además hace gráficas de rosco para analizar mejor dicha información. Más adelante, en el siguiente bloque del trabajo, se verá un ejercicio práctico con este programa para comprobar su funcionamiento y qué resultados se pueden obtener y cómo los presenta.

## 6. EXPERIMENTOS

### 6.1. RAPIDMINER

La primera herramienta de esta parte práctica será RapidMiner. Con ella analizaremos en Twitter el *hashtag* #Elecciones10N con motivo de las elecciones generales celebradas en España el 10 de noviembre y el cual fue muy utilizado tanto el pasado 4 de noviembre en el debate electoral que realizaron los representantes de cada uno de los partidos principales y que fue retransmitido por TVE en directo, como el mismo día de la celebración de las elecciones. En Twitter fue *trending topic* en España durante todo el día destacando en las horas del recuento de votos y una vez publicados los resultados. Durante los días siguientes se ha seguido utilizando el *hashtag* pero en menor medida y solamente para casos específicos.

Una vez explicadas las nociones básicas de lo que se va a analizar, se hará una pequeña toma de contacto con el programa y en los puntos siguientes se expondrán cuáles son los operadores necesarios para realizar las operaciones y ejecutar el análisis deseado.

Al abrir el programa para realizar el proyecto nuevo, se verá la interfaz principal que incluye una pantalla central de procesos más grande que el resto, un menú de repositorio en la esquina superior izquierda (desde el que se podrá acceder a las diferentes carpetas de almacenamiento del proceso), un menú de operadores debajo de este (donde se agregaran todos los que necesitemos para analizar las diferentes redes sociales), una pantalla de parámetros en el lado derecho y debajo de esta, una pestaña de ayuda donde se definen los diferentes operadores que usemos y su utilización, algo que puede ser de gran ayuda para los principiantes (Fig. 3).

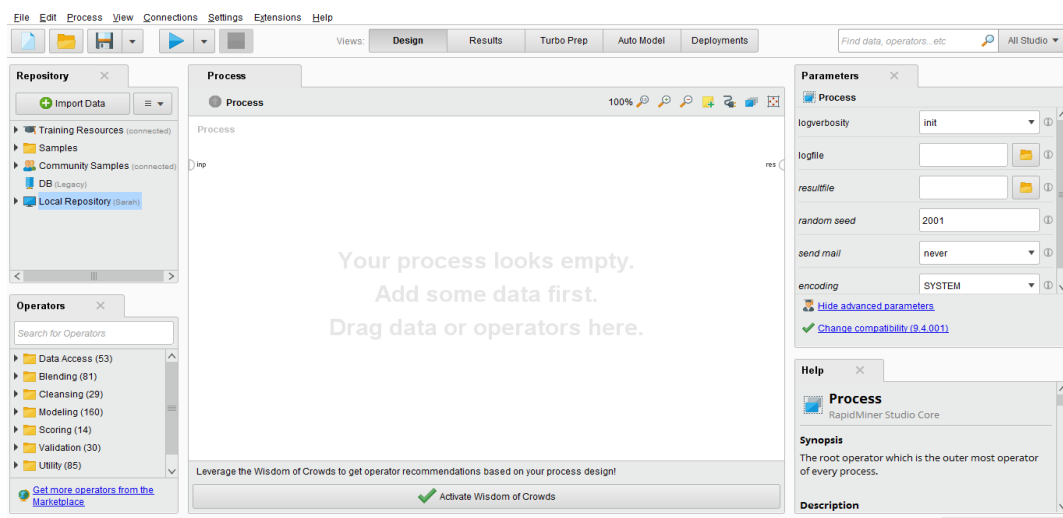


Fig. 3. Pantalla principal del programa RapidMiner

## 6.1.1. OPERADORES NECESARIOS

### 6.1.1.1. OPERADOR *TWITTER*

Este es el operador más importante ya que el análisis a realizar es de los usuarios de esta red social. Para incluirlo en el proyecto tan solo habrá que realizar la búsqueda “Twitter” en la pantalla de operadores y seleccionar “*Search Twitter*”. Una vez encontrado, hay que arrastrarlo hasta la pantalla de procesos. Antes de programarlo se debe crear una conexión a una cuenta de Twitter. Para ello, se seleccionará “*Manage Connection (Legacy)*” en la pestaña desplegable superior “*Connections – Legacy Connections*”, se añadirá una nueva conexión y se rellenarán los campos requeridos (Fig. 4).

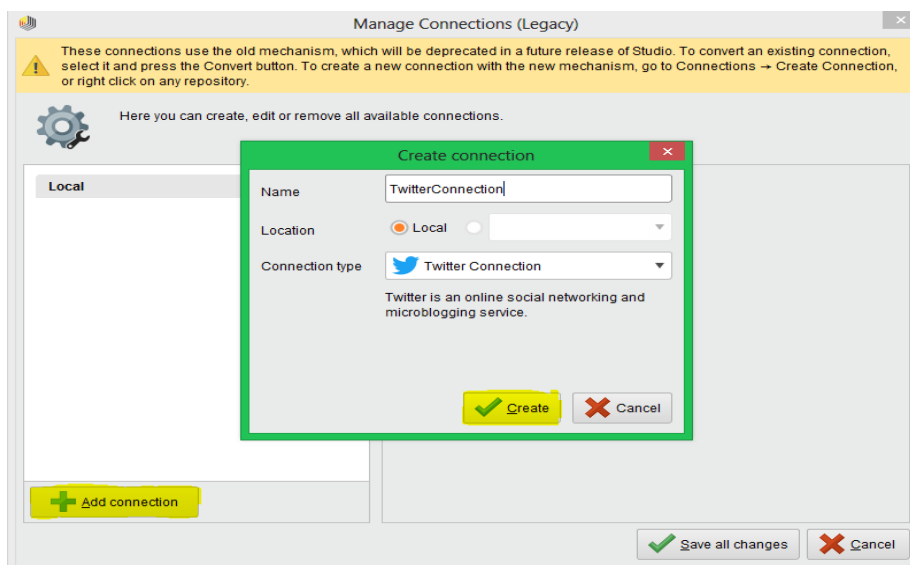


Fig. 4. Creación de la conexión del operador con una cuenta de Twitter

Al crear la conexión, se debe autorizar a RapidMiner a utilizar la cuenta de Twitter conectada. Para ello, se clicará en el recuadro pequeño que aparece en la parte derecha del recuadro de código “*Access Token*” (1). Tras esto, pinchando en “*Request Access Token*” (2), pedirá que se autorice a la aplicación para utilizar la cuenta de Twitter y generará un código que copiaremos en el recuadro correspondiente de la autorización (3). El resultado será un código que se añadirá automáticamente a “*Access Token*”. Pinchando en “*Convert*” (4) se abrirá la ventana de creación de la conexión y al aceptar, el código automático completo aparecerá en el campo correspondiente (5) (Fig. 5). Para que se guarde correctamente la conexión, se deberá pinchar en el botón “*Save all changes*” antes de pasar a los siguientes pasos.

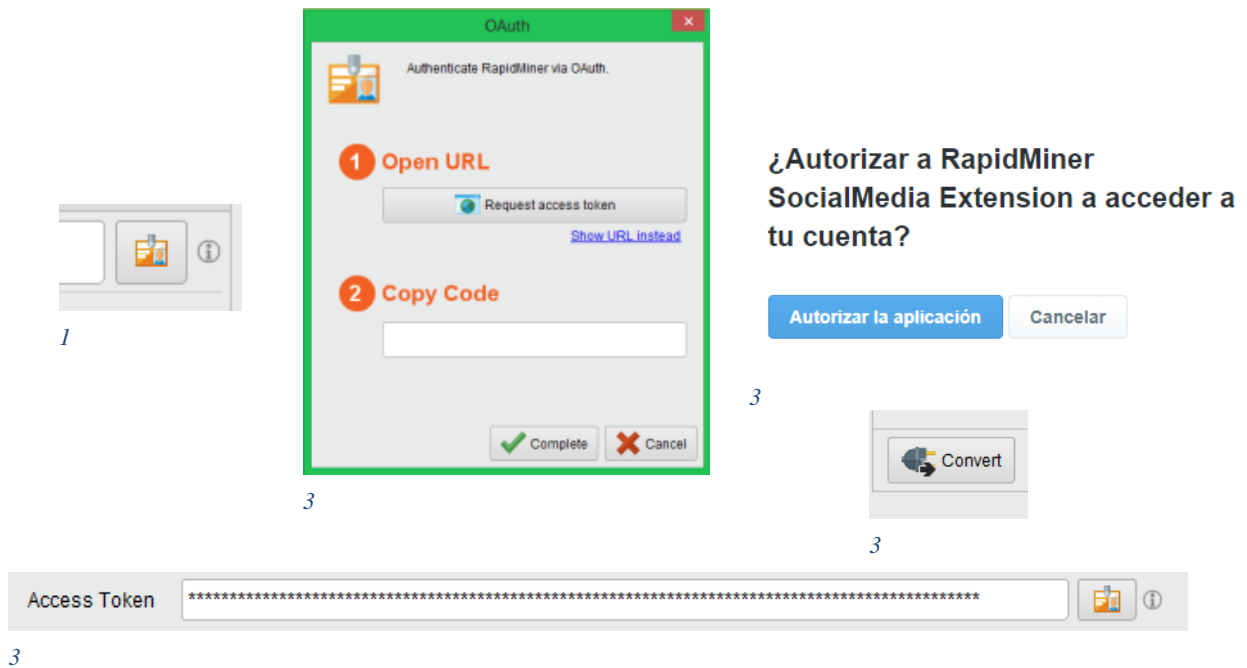


Fig. 5. Pasos a realizar para autorizar la conexión con Twitter

Una vez hecha la conexión se ajustarán los parámetros para la búsqueda y análisis que se quiere realizar. Si se pincha en el icono de Twitter de la pantalla de procesos aparecerán a la derecha los parámetros a completar para esta búsqueda (Fig. 6).

- **Connection entry:** aquí se incluirá la conexión que se ha realizado anteriormente.
- **Query:** es el campo en el que se escribirá el usuario (@) o el *hashtag* (#) a analizar. Como no queremos que lea los retweets, se añade “-rt” tras el *hashtag*.
- **Result type:** se podrá elegir si se quiere que el análisis se haga entre lo más reciente, lo más popular o una combinación de ambos.
- **Limit:** en este apartado se podrá delimitar la cantidad de tweets a analizar.
- **Language:** se podrá escoger el lenguaje en el que se han escrito los *tweets* siguiendo la norma ISO 639-1.

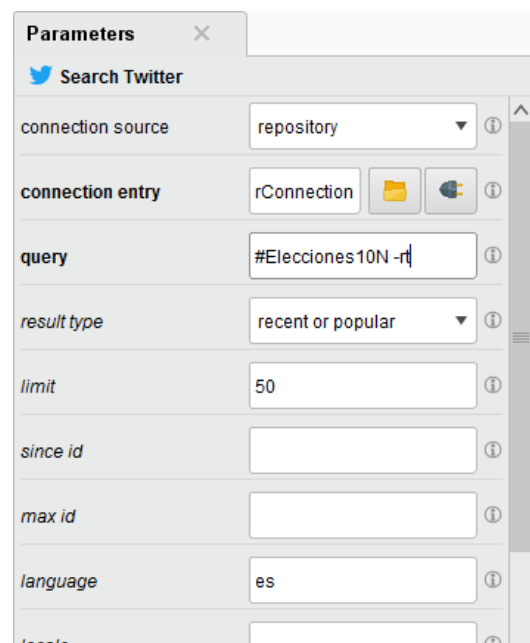


Fig. 6. Parámetros para la búsqueda y análisis

### 6.1.1.2. OPERADOR ANALYZE SENTIMENT

Una vez incluido el operador de Twitter, se debe añadir el operador que se encargará de extraer los tweets y hacer el análisis de sentimientos de los mismos. Para ello se usará el *Text Analysis API Documentation* de AYLIEN. En la misma página web de esta se puede encontrar un tutorial paso a paso y muy explicativo de instalación de la extensión en RapidMiner y como se puede hacer el análisis de sentimientos, algo que se verá más adelante. Este apartado se centrará en la instalación de la extensión y su puesta en funcionamiento.

La extensión ya se encuentra en el *software* por lo que tan solo habrá que encontrarla. Para ello, se hace la búsqueda “*marketplace*” en la esquina superior derecha de la pantalla principal y, una vez dentro, se buscará “Ayllen” para encontrar así la extensión deseada. El primero que aparecerá será el que se utilizará para realizar con éxito el análisis de sentimientos (Fig. 7).

Una vez instalado, el operador se podrá encontrar en la pantalla de operadores tal y como se hizo anteriormente con el de Twitter. Como se puede comprobar no solo se ha instalado uno, pero el que se utilizará en este caso es el operador llamado “*Analyze Sentiment*”. Al igual que antes, se arrastra a la pantalla de procesos y se hará una nueva conexión y configuración.

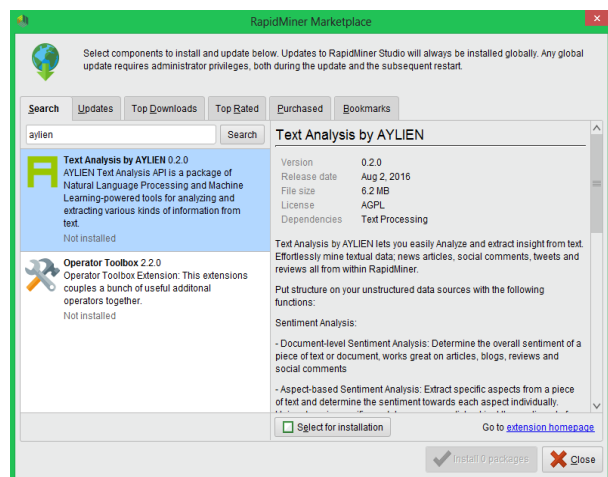


Fig. 7. Operador AYLIEN

En este caso la creación de la conexión será un poco diferente ya que se deben obtener unas credenciales específicas para conectarnos a la API de AYLIEN. Se necesita una APP ID y una clave de API para conectar con AYLIEN para lo que habrá que registrarse en el siguiente enlace. Tras realizar las confirmaciones de correo electrónico correspondientes a este tipo de registros y poder acceder al tablero de la API, se verán los datos que necesitamos. Seguidamente se crea la nueva conexión siguiendo los pasos anteriores, se añaden las credenciales y se guardan los cambios. Al igual que antes, se configuran los parámetros para completar la conexión del API (Fig. 8). Esto realiza un análisis básico de los tweets y determina la polaridad y subjetividad.

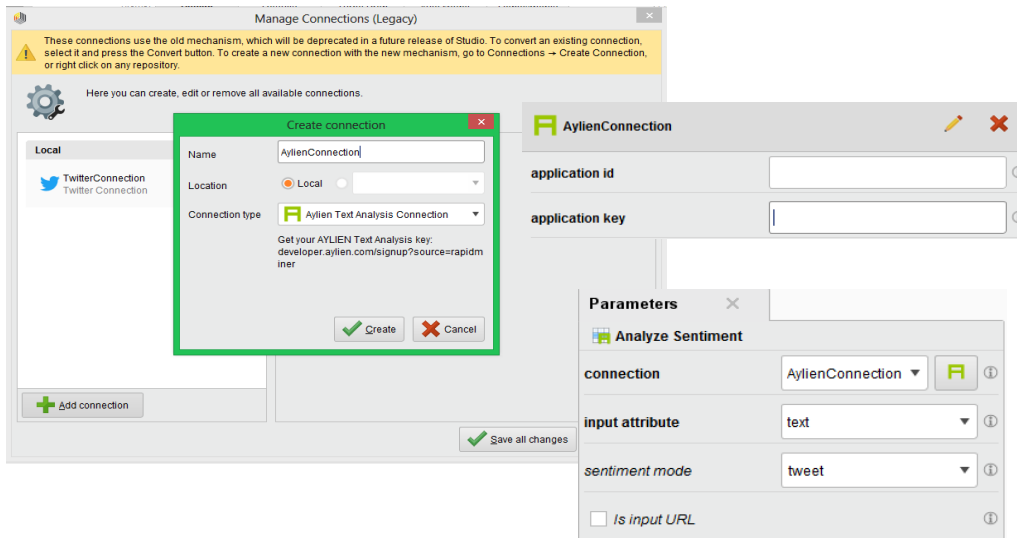


Fig. 8. Creación de la conexión AYLIEN

Tras toda esta primera parte se puede hacer un análisis básico de los tweets determinado su polaridad, pero también se quieren categorizar. En los siguientes apartados se verá cómo completar el análisis.

#### 6.1.1.3. OPERADOR *DATA TO DOCUMENT* (Fig. 9)

Este genera documentos a partir de datos para facilitar su categorización. Su configuración es muy sencilla pues apenas se necesitan los parámetros que se obtienen por defecto.

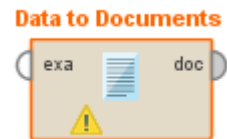


Fig. 9. Operador *Data to Document*

#### 6.1.1.4. OPERADOR *CATEGORIZE (DOCUMENT)* (Fig. 10)

Este operador ayuda a clasificar semánticamente el texto de acuerdo a la taxonomía especificada que, en este caso, será la misma que se utiliza en la explicación-tutorial que ofrece AYLIEN en su página web: IAB QAG *taxonomy*. Al igual que antes, los parámetros por defecto son suficientes para su correcto funcionamiento.

Este opera por una conexión al servidor AYLIEN por lo que será necesario conectarse a la conexión creada anteriormente al empezar a usar dicha extensión.

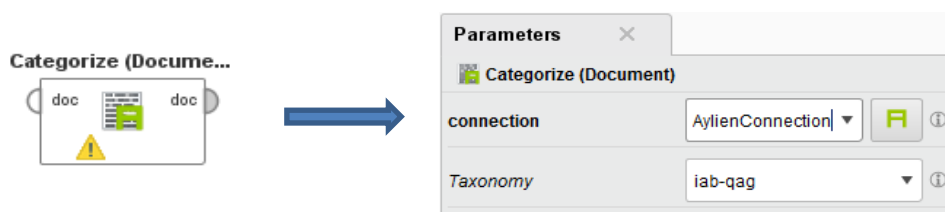


Fig. 10. Operador *Categorize* y parámetros

### 6.1.1.5. OPERADOR *DOCUMENT TO DATA* (Fig. 11)

Este bloque extrae y genera datos a partir de documentos, proceso inverso al realizado con el operador *data to document*. Así se creará un conjunto de datos como resultado de los documentos categorizados, lo que hará más fácil la visualización y comprensión de los resultados finales que se obtengan. Igualmente, los parámetros por defecto son suficientes para la configuración de este operador.

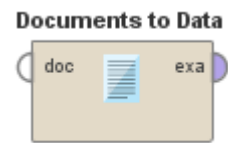


Fig. 11. Operador *Document to Data*

### 6.1.2. RESULTADOS

Una vez realizados todos estos pasos, se obtendrá en la pantalla de procesos el árbol completo del proyecto actual (Fig. 12). Uniendo cada uno de los parámetros, y el último de ellos al punto final “res” (resultados), y pinchando en el botón “play”, el programa comenzará a cargar cada bloque para obtener así los resultados deseados.

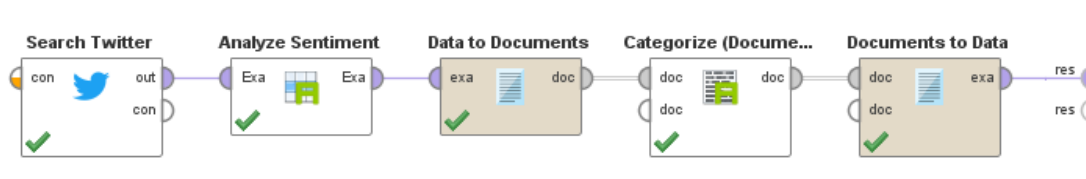


Fig. 12. Árbol de procesos definitivo

Para crear un gráfico en RapidMiner pinchamos en “Visualizations”. Por defecto aparecerá un diagrama de barras, pero en este caso se seleccionará el estilo “Pie” para crear un diagrama diferenciado por sectores y se configurará tal y como en la figura 13. El diagrama resultante será el mostrado en la figura 14.

Los resultados más interesantes en este proyecto son los relativos a la polaridad ya que lo que se desea mostrar es un ejemplo de análisis de sentimientos con el *software* RapidMiner.

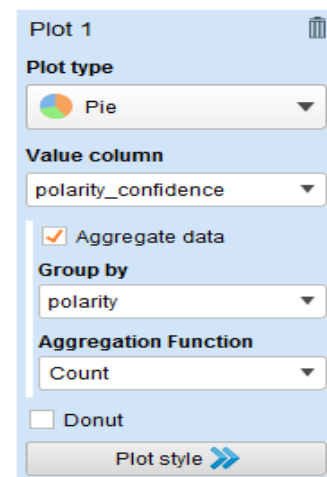
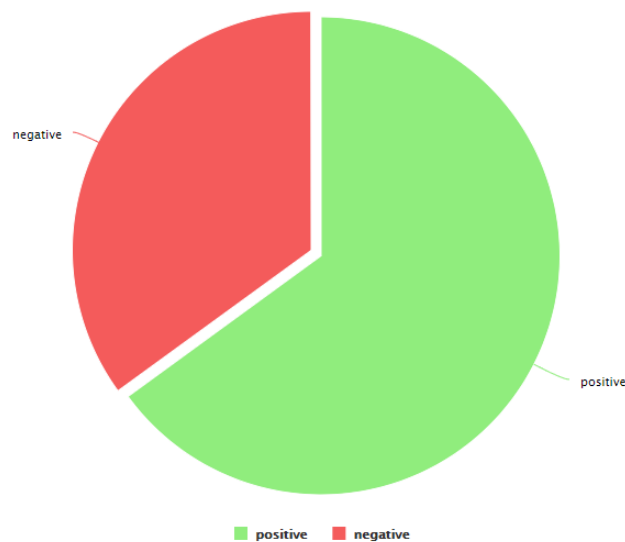


Fig. 13. Configuración del diagrama

En total se evaluaron 100 tweets de los cuales 65 tienen un valor positivo y los 35 restantes negativo. Esto significa que de los usuarios estudiados que twitearon sobre las elecciones generales celebradas en España el día 10 de noviembre incluían en estos mensajes connotaciones positivas. Sin embargo, conociendo el recibimiento que ha tenido este tema en



la sociedad española, se podría baremar la posibilidad de un error en el análisis o una falta de comprensión por parte del programa. Esto es de lo que se habló en el primer capítulo de este proyecto. Los programas de análisis de sentimientos actuales no son capaces de entender la ironía y el sarcasmo, sentimientos que pueden ser muy comunes en los comentarios sobre este tipo de temas políticos y sociales. De hecho, si se leen algunos de los tweets que se han analizado, se puede identificar que es un fallo de comprensión de la máquina y el *software* ya que algunos tweets que claramente tienen una polaridad negativa, están clasificados como positivos y viceversa. Pero no se le dará mucha importancia en este caso al resultado pues la finalidad del trabajo es explicar el funcionamiento de algunas herramientas y esclarecer lo básico sobre el análisis de sentimientos. Si el objetivo del trabajo fuese hacer un análisis concreto de sentimientos sobre ciertos temas, se deberían estudiar cuales son los factores, e intentar ponerles solución, por las que los resultados obtenidos son tan dispares e incorrectos.



*Fig. 14. Diagrama de resultados obtenido al ejecutar el proceso en el software*

## **6.2. BRAND24**

La segunda herramienta a utilizar será Brand24, con la que se analizará el mismo *hashtag* que con la herramienta anterior, RapidMiner, y se compararán de esta forma los resultados que se obtienen y la forma de mostrar el análisis. Esto, además, permitirá hacer una comparativa, que se verá en el siguiente apartado, de las funcionalidades de ambas herramientas y su facilidad para el uso del usuario inicial.

Como se ha dicho en el apartado de herramientas, Brand24 ofrece una prueba gratis de 14 días, menos que la mayoría de programas, pero para su utilización tan solo habrá que proporcionar un correo electrónico y una contraseña.

Una vez hecho esto, inmediatamente aparece una pantalla de búsqueda, en la que se debe introducir el tema, entidad, marca o, en este caso, *hashtag*, a tratar. Se pueden incluir incluso varias entradas. Se pincha en “*Create Project*” y se inicia el proceso.

La pantalla principal muestra ya directamente el análisis completo con una serie de datos que se explicarán en el siguiente punto. Este análisis se encuentra en el apartado “*Mentions*” del menú que se encuentra en la parte lateral izquierda del escritorio. En este se incluye todos los proyectos creados (cuya monitorización será única y diferente) y una multitud de funcionalidades, características y ajustes para cada uno de los proyectos existentes (fig. 15). Se explicaran cada una de estas secciones por separado para mejor comprensión.

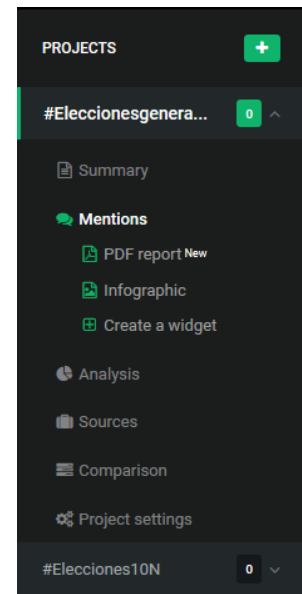


Fig. 15. Menú de ajustes del proyecto.

### 6.2.1. MENTIONS: escritorio principal del análisis

Lo primero que aparece en la ventana es una gráfica que compra en dos líneas el número de menciones y el alcance mediático de los tweets. En la esquina derecha hay un filtro de fechas donde se podrá seleccionar el rango temporal que se quiere analizar o del que se quiere extraer el análisis. Además, en la esquina contraria, si se pincha en el corazón, se podrá acceder a una gráfica de análisis de sentimientos (fig. 16).

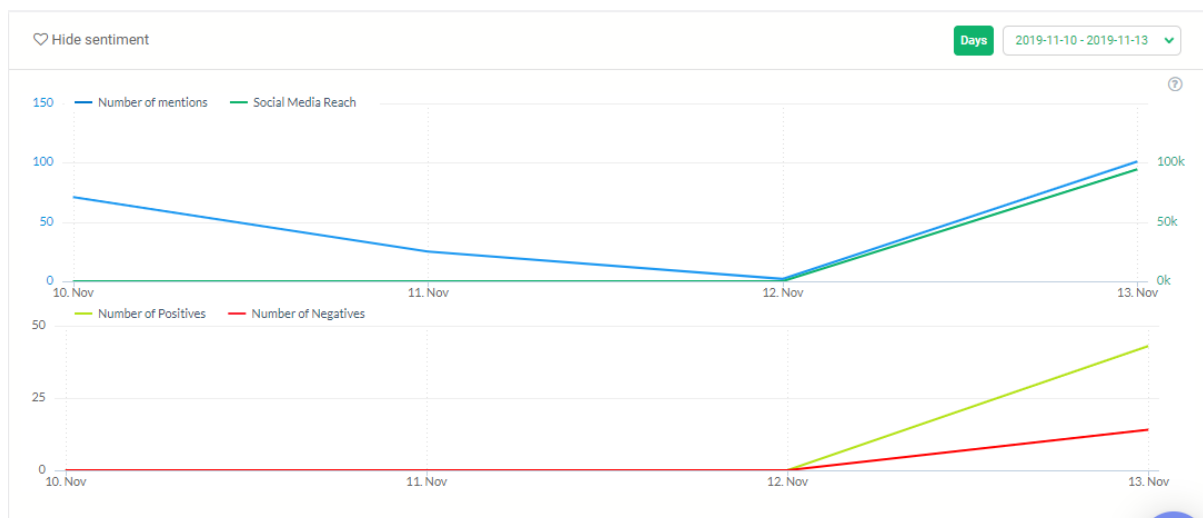


Fig. 16. Gráficas de número de menciones, alcance mediático y análisis de sentimientos.

Más abajo se pueden encontrar una serie de recursos de los que se puede extraer información útil como son la red social Twitter, noticias, vídeos de YouTube, Foros... Bajo estos, una lista de lo que se ha recuperado. En este caso, como se ha seleccionado la opción de Twitter, se leerá una lista de los tweets más recientes publicados con la búsqueda realizada en el proyecto. Como se observa, hay una pestaña, que en la figura 17 está en verde y en la que se encuentra escrito “Positive”, que mostrará el sentimiento detectado por el programa en el tweet. Algo muy interesante es que se puede cambiar la polaridad si se considera que el sentimiento del mensaje es otro diferente al que ya se encuentra marcado. Como siempre se podrá elegir entre positivo, negativo o neutro. Además, se puede ver mucha más información que puede ser útil como son el marcador de influencia del tweet, los “me gusta” o “retweet” que tiene, la fecha de publicación, etc. Algo muy curioso es el cambio masivo que se puede hacer. Pinchando en el cuadradito blanco que aparece en las esquinas inferiores derechas de los tweets se pueden seleccionar para, más tarde, en el final del escritorio, hacer cambios en todos los tweets marcados.

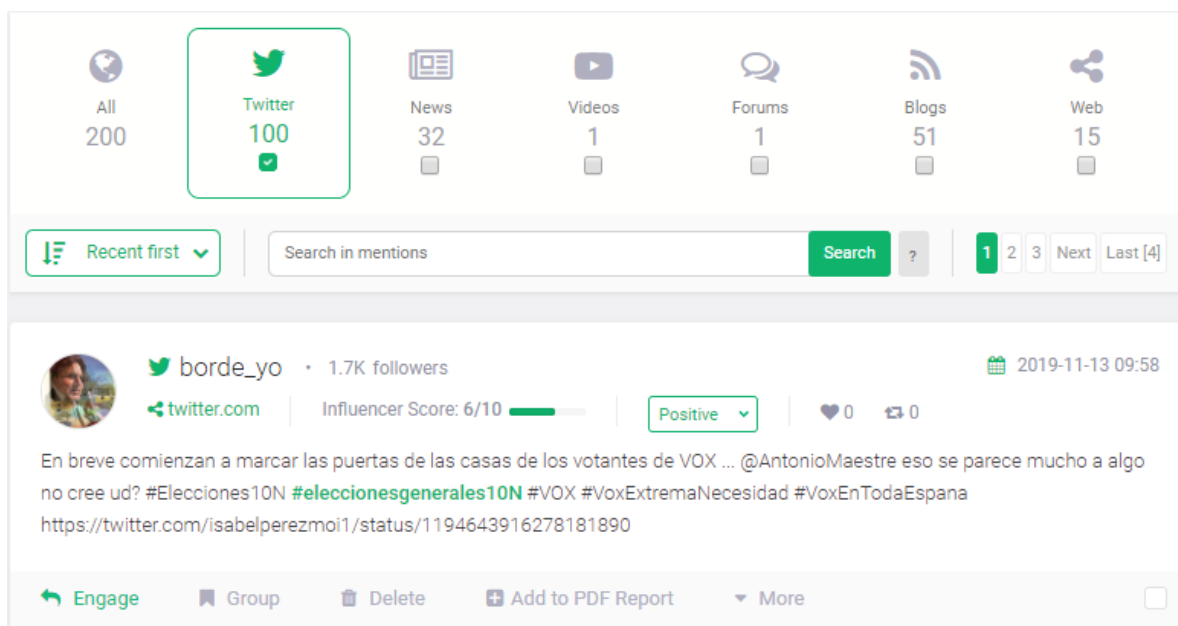


Fig. 17. Ventana de plataformas a analizar y tweets analizados.

Si desde esta misma ventana se accede al menú lateral izquierdo ya nombrado se puede descargar un PDF con datos obtenidos en el análisis. Antes de la descarga se deben activar una serie de botones para indicar que información queremos que incluya el documento y una vez hecho esto se podrá descargar pinchando en “Generate report” (fig. 18). Este documento

es un análisis muy completo con toda la información que ofrece el análisis bastante bien expuesta evitando así los posibles problemas de confusión o falta de comprensión.

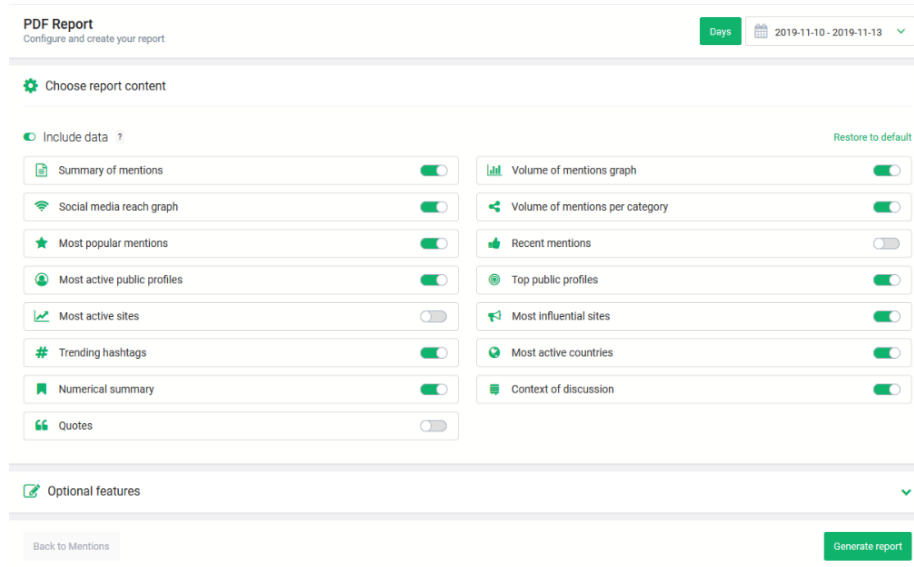


Fig. 18. Botones para incluir información en el análisis descargable en formato PDF.

La opción “Infografic” genera un resumen de los resultados, pero de forma gráfica, lo que hace que sea muchísimo más fácil de comprender por el usuario al ser más visual (fig. 19). Este resumen incluye, de nuevo, de qué fuentes se recoge la información y cuáles son las más activas, un gráfico de líneas y otro numérico del análisis de sentimientos, los usuarios más influyentes y activos...

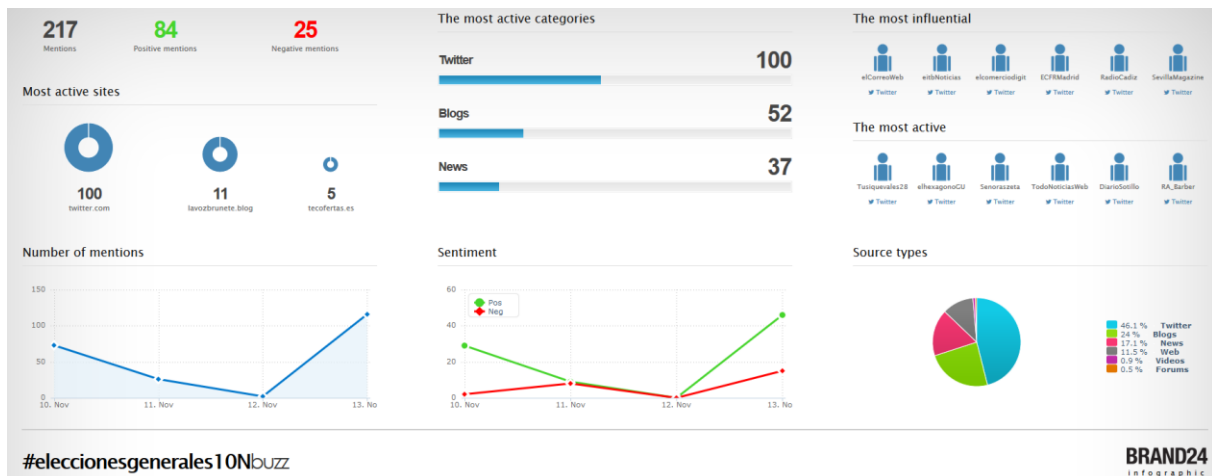


Fig. 19. Infografía del análisis.

A toda esta información a la que se accede, o que se resume tanto en el PDF como en la infografía, se puede acceder desde la misma herramienta de forma más extensa. Se debe recordar que ambos archivos recuperados anteriormente son resúmenes.

### 6.2.2. SUMMARY

La opción “*Summary*” ofrece un resumen del análisis incluyendo las menciones más populares y más recientes, gráficas y valores numéricos de menciones, alcance mediático y análisis de sentimientos, los perfiles más influyentes, estadísticas, los países más activos en el tema, etc.

### 6.2.3. ANALYSIS

Otra opción que será de mucha utilidad en el tipo de análisis que se está realizando en este caso es la de “*Analysis*”. En esta opción encontramos todo lo que se podía encontrar en los resúmenes, pero de forma mucho más detallada. Esta opción incluye, por ejemplo, una cuadrícula con los resultados numéricos del análisis, una relación de los *hashtags* más utilizados por todos los usuarios de la red social, los sitios web más influyentes en la sociedad o una nube de palabras clave que pueden ayudar al usuario para futuros análisis sobre el mismo tema (fig. 20).

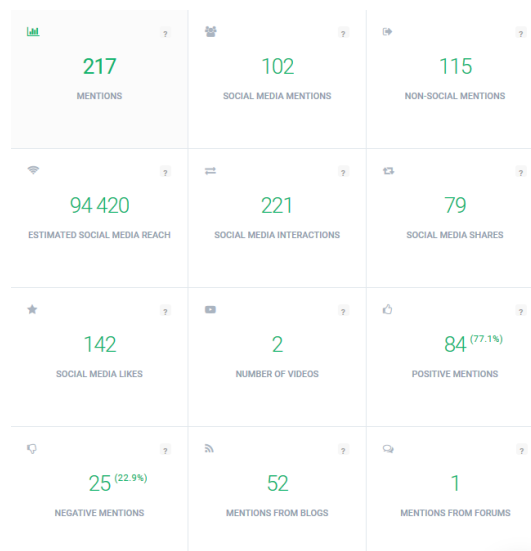


Fig. 20. Análisis numérico.

### 6.2.4. COMPARISON

La opción “*Comparison*” puede resultar muy útil si el análisis realizado es sobre una marca o entidad y se quiere comparar esta con otra del mismo tipo. Por ejemplo, si se comparan las marcas Coca-Cola y Pepsi en la web se obtienen los siguientes resultados (fig. 21). Lo primero que aparece es una gráfica comparativa de las menciones en los últimos 30 días que han recibido ambas marcas. Como se puede ver, Coca Cola suele ser la marca más mencionada en las redes. Debajo de esta, cuatro gráficos con forma de rosco y una serie de

marcas numéricas, compararán ambas marcas según las fuentes de donde se han obtenido los resultados, las menciones que tienen y el porcentaje de sentimientos positivos o negativos que se obtienen del análisis.



Fig. 21. Comparación de las marcas Pepsi y Coca Cola.

### 6.2.5. PROJECT SETTINGS

En este apartado se pueden configurar algunos ajustes para indagar o acotar la información que se quiere obtener. Aparecen cinco opciones diferentes que se pueden configurar:

- **Keywords:** se podrán añadir más palabras clave al proyecto, seleccionar el idioma en el que se quieren recuperar los mensajes e incluso añadir palabras que deben aparecer de forma obligatoria en el mensaje o palabras que se quieren excluir del mismo.
- **Sources:** en este apartado se podrá seleccionar la fuente de la que se quieren extraer los datos para que no aparezcan resultados de otras redes sociales o sitios web. Se podrá incluso excluir algunos sitios de los que no se quiera obtener datos.
- **Notifications:** desde aquí se podrá crear una alarma de notificación, desde a tiempo a real como a cada mes, para que envíe nuevos datos al correo electrónico que se seleccione.

- **Reports:** es otro tipo de alarma que enviará al correo electrónico un informe con los datos nuevos que han surgido o tenido lugar durante la ausencia del usuario en el escritorio del programa.

Tal y como RapidMiner, y como se comprobó en la utilización de este programa, Brand24 no termina de ofrecer resultados absolutamente fiables pues, al leer muchos de los comentarios, se descubre como la clasificación de sentimientos es errónea al entendimiento humano. Esto hace que los resultados obtenidos no sean del todo fiables, aunque si pueden dar una ligera idea de lo que ocurre en la web y cuál es la opinión general de la gente. Probablemente esto se deba también al hecho de que en este trabajo se ha usado una versión gratuita de demostración a la que probablemente le falten algunas de las funciones que si tendría en la versión de pago. Pero al fin y al cabo en este trabajo lo que se desea es mostrar cómo pueden funcionar este tipo de herramientas, para qué y qué información aportan sobre el análisis de sentimientos.

## 7. CONCLUSIONES

A día de hoy, el análisis de sentimientos sigue sin aportar el 100% de la precisión que se requiere de tal trabajo para hacerlo absolutamente fiable. Pero el trabajo y la investigación continuos en este campo hacen que poco a poco existan muchas más herramientas que puedan realizar este trabajo y hacerlo cada vez más eficaz. Desde el aprendizaje automático hasta las técnicas basadas en diccionarios son objeto de la continua investigación para salvar las limitaciones del análisis de sentimientos y conseguir así dar un gran paso en el crecimiento de la inteligencia artificial.

Ya se habla de futuras máquinas y robots que, a través de este tipo de técnicas, sean capaces de comunicarse con los seres humanos y reaccionar a los sentimientos de los mismos, no solo entendiendo el significado de las palabras sino también las intenciones que hay detrás de ellas. Pero actualmente se usa este tipo de tarea, sobretodo, para realizar análisis de marketing, comparaciones entre marcas, para conocer la reacción de los consumidores ante un nuevo producto o para prever los resultados posibles en unas elecciones. El desarrollo de este tipo de técnicas es muy rápido en cuanto a aparición de nuevas herramientas capaces de realizar dicha tarea y muy lento a la vez, pues su innovación y desarrollo conlleva un estudio muy profundo del funcionamiento del ordenador, cómo recoge el lenguaje natural usado por

el ser humano y como puede conseguir entender dicha información para aportar datos sobre la misma.

En el apartado de herramientas se comprobado la dificultad que existe para entender el dicho lenguaje natural del que se habla, como unos comentarios extraídos de Twitter son positivos o negativos a nuestra visión como seres humanos y como sin embargo el *software* lo entendió de una forma totalmente diferente. Esto hace que los resultados obtenidos no sean del todo fiables. Sin embargo, se debe tener en cuenta que ambas herramientas utilizadas han sido a través de una prueba gratuita y que este trabajo es meramente académico. En contraposición, la mayoría de herramientas existentes que utilizan las grandes empresas o incluso los gobiernos para estudiar qué se dice de ellos en la web, son de pago, tienen muchísimas funcionalidades y están gestionadas por profesionales de la informática e inteligencia artificial que se encargan de que los resultados obtenidos sean lo más fiable posible.

De las dos herramientas usadas para este trabajo, para mí la más completa es Brand24, pues es muy intuitiva, se puede acceder a los análisis y estadísticas desde cualquier lugar pues se encuentra almacenada en la nube y dichos análisis son muy completos. Sin embargo, la recuperación de información es un poco difícil de entender pues, personalmente he encontrado errores de congruencia si se filtraba la búsqueda por un tiempo limitado. RapidMiner, en la otra mano, también ofrece un análisis muy completo pero esta herramienta si tiene una versión a descargar e instalar en el ordenador y su utilización no es para nada intuitiva. Tanto la utilización de los operadores como la instalación de la API no son nada sencillos de utilizar por una persona que no tenga experiencia. Además, las guías formativas que se encuentran en AYLIEN son mucho más extensas y útiles que las utilizadas en RapidMiner.

Aun así, y tras mi opinión personal, ambas pueden dar una idea bastante clara de cómo funcionan este tipo de herramientas y qué pueden hacer para extraer la información que, al final, interesa para poder reaccionar ante las actitudes de los usuarios: los sentimientos no explícitos en el mensaje o publicación de los usuarios.



## BIBLIOGRAFÍA

- Arcila-Calderón, C., Ortega-Mohedano, F., Jiménez-Amores, J. y Trullenque, S. (2017): “Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tweets basada en aprendizaje automático”, *El profesional de la información*, 26(5), pp. 973-982. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=6109608> [Consulta: 1 de octubre de 2019]
- Baviera, T. (2017): “Técnicas para el análisis del sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength”, *Revista de Comunicación Digital DÍGITOS*, 1(3), pp. 33-50. Disponible en: <https://revistadigitos.com/index.php/digitos/article/view/74/39> [Consulta: 1 de octubre de 2019]
- BD Guidance (2016): *Primer Taller de Análisis de Sentimiento en Twitter con R*. [Video online]. Disponible en: <https://www.youtube.com/watch?v=nOIZnYLIPBo> [Consulta: 2 de septiembre de 2019]
- Bing, L. (2006): “Opinion Mining and Sentiment Analysis”, en *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Disponible en: DOI 10.1007/978-3-642-19460-3 [Consulta: 18 de septiembre de 2019]
- Bravo-Marquez, F., Mendoza, M. y Poblete, B. (2014): “Meta-level sentiment models for big data social data analysis”, *Knowledge-Based Systems*, 69 (1), pp. 86-99. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0950705114002068> [Consulta: 13 de septiembre de 2019]
- Carpenter, M. (2018): “Las mejores herramientas para analizar el sentimiento de marca”. *Talkwalker*, 24 de julio. Disponible en: <https://www.talkwalker.com/es/blog/las-mejores-herramientas-analizar-sentimiento> [Consulta: 18 de octubre de 2019]
- González, P. (2011): “Social Analytics: Radian6 vs Socialmention”. *#Analiticaweb: Reflexiones sobre madurez digital, datos y tecnología*, 8 de junio. Disponible en: <https://www.analiticaweb.es/social-analytics-radian6-vs-socialmention/> [Consulta: 18 de octubre de 2019]

- Hootsuite Inc. (2019): *Hootsuite Insights: Mejor escucha social. Decisiones empresariales más inteligentes*. Disponible en: <https://hootsuite.com/es/productos/insights#> [Consulta: 23 de octubre de 2019]
- Li, Q. (2015): *Examining the accuracy of sentiment analysis by Brand monitoring companies*. 5th IBA Bachelor Thesis Conference. University of Twente. Disponible en: <https://www.semanticscholar.org/paper/Examining-the-accuracy-of-sentiment-analysis-by-Li/ae82ed8aa774bebd137232605b006292d195583f> [Consulta: 12 de noviembre de 2019]
- MarTech Forum, S.L.U. (2019): *Brand24*. Disponible en: <https://www.martechforum.com/herramienta/brand24/> [Consulta: 30 de octubre de 2019]
- Medhat, W., Hassan, A. y Korashy, H. (2014): “Sentiment analysis algorithms and applications: A survey”, *Ain Shams Engineering Journal*, 5(4), pp. 1093-1113. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2090447914000550> [Consulta: 18 de septiembre de 2019]
- Panico, C. (2018): *La eficacia del análisis de sentimientos para la empresa: el caso de estudio Dell Technologies Inc*. Trabajo de Fin de Grado. Universidad Complutense de Madrid. Disponible en: [https://www.ucm.es/data/cont/docs/758-2019-01-04-TFG\\_Panico\\_Chiera\\_TFG.pdf](https://www.ucm.es/data/cont/docs/758-2019-01-04-TFG_Panico_Chiera_TFG.pdf) [Consulta: 05 de noviembre de 2019]
- RapidMiner, Inc. (2019): *RapidMiner, Lightning Fast Business Impact with AI*. Disponible en: <https://rapidminer.com/> [Consulta: 18 de octubre de 2019]
- Semantic Engines LLC. (2010-2019): *Opinion Crawl: Online sentiment analysis for current events, companies, products, and people!*. Disponible en: <http://www.opinioncrawl.com/> [Consulta: 18 de octubre de 2019]
- Serrano-Guerrero, J., Olivas, J.A., Romero, F.P., Herrera-Viedma, E. (2015): “Sentiment analysis: A review and comparative analysis of web services”, *Information Sciences*, 311, pp. 18-38. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0020025515002054?via%3Dihub> [Consulta: 12 de septiembre de 2019]
- Talkwalker Inc. (2019): *Talkwalker: El impacto empieza aquí*. Disponible en: <https://www.talkwalker.com/es> [Consulta: 18 de octubre de 2019]

- Tripathi, G. y Naganna, S. (2015): “Feature selection and classification approach for sentiment analysis”, *Machine Learning and Applications: An International Journal (MLAIJ)*, vol. 2 (no2). Disponible en: DOI 10.5121/mlaij.2015.2201 [Consulta: 12 de noviembre de 2019]
- Velasco, J.J. (2010): “Cinco herramientas para analizar los sentimientos de los tweets”. *Hipertextual*, 3 de diciembre. Disponible en: <https://hipertextual.com/archivo/2010/12/cinco-herramientas-para-analizar-los-sentimientos-de-los-tweets/> [Consulta: 18 de octubre de 2019]
- Wilson, T., Wiebe, J. y Hoffmann, P. (2005): “Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis”, *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*. Vancouver, octubre de 2005. Association for Computational Linguistics, pp. 347-354