



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Politécnica Superior de Gandia

Análisis de semifinales y finales de los playoffs de la NBA
2021-2022 a través de Twitter

Trabajo Fin de Máster

Máster Universitario en Comunicación Transmedia

AUTOR/A: Moreno Salvador, Rubén

Tutor/a: Rodríguez Rodríguez, Alejandro

CURSO ACADÉMICO: 2021/2022

Resumen

En este trabajo de investigación se han aplicado técnicas aprendidas durante el curso académico. Concretamente el análisis de sentimiento de textos en Twitter y la *clusterización* de usuarios. Para esto se realizó una gran base de datos (formada por cuatro pequeñas bases de datos) de tweets descargados, cerca de 10.000, gracias a una extensión de Google Sheets. Después de haber limpiado esta base de datos ha sido tratada con el *software* 'Azure' (análisis de sentimiento) y 'XLStat' (*clusterización*). Además, para tener una segunda opinión se ha hecho un análisis de sentimiento manual con una de las bases de datos para contrastar los resultados con los de 'Azure'. Todo este procedimiento se ha hecho con el objetivo principal de detectar errores en las técnicas actuales de estos análisis y, a su vez, proponer mejores para subsanarlos.

Palabras clave

Twitter, redes sociales, usuarios, análisis de sentimiento, *clustering*

Abstract

In this research work, techniques learned during the academic year have been applied. Specifically, the sentiment analysis of texts on Twitter and the clustering of users. For this, a large database (formed by four small databases) of downloaded tweets was made, about 10,000, thanks to a Google Sheets extension. After having cleaned this database, it has been treated with the software 'Azure' (sentiment analysis) and 'XLStat' (clustering). In addition, to have a second opinion, a manual sentiment analysis has been carried out with one of the databases to compare the results with those of 'Azure'. All this procedure has been done with the main objective of detecting errors in the current techniques of these analyzes and, in turn, proposing better ones to correct them.

Keywords

Twitter, social media, users, sentiment analysis, clustering

Índice de contenidos

1. Introducción.....	pág.9
2. Motivación.....	pág.11
3. Estructura.....	pág.12
4. Objetivos.....	pág.13
5. Metodología.....	pág.14
5.1. Generar la base de datos.....	pág.14
5.2. Análisis de sentimiento automático.....	pág.16
5.3. Análisis de sentimiento manual.....	pág.18
5.4. Análisis de clusters.....	pág.18
6. Estado de la cuestión.....	pág.19
6.1. Análisis de sentimiento.....	pág.19
¿Cómo surge?.....	pág.19
¿Cómo lo conocemos a día de hoy?.....	pág.20
¿Qué es el Procesamiento del Lenguaje Natural (PLN)?.....	pág.21
¿Qué es el análisis de sentimiento basado en diccionarios?.....	pág.22
¿Qué aplicaciones disponemos a día de hoy para el análisis de sentimiento?.....	pág.22
<i>Softwares</i> basados en el PLN.....	pág.22
<i>Softwares</i> basados en diccionarios.....	pág.24
6.2. Clustering.....	pág.24
¿Cómo surge y qué es el <i>clustering</i> ?.....	pág.24
<i>Atribus, Social Listening</i>	pág.25

Caso nº1. Nuevas formas de consumir café	pág.27
Caso nº2. Valencia CF, HBOmax y The Batman	pág.28
7. Resultados y discusión	pág.30
7.1. Análisis de sentimiento automático	pág.30
7.2. Análisis de sentimiento manual	pág.35
7.3. Análisis de <i>clusters</i>	pág.39
8. Conclusiones	pág.46
8.1. Conclusiones del análisis de sentimiento	pág.46
8.2. Conclusiones del análisis de <i>clusters</i>	pág.46
9. Relación del trabajo desarrollado con los estudios cursados	pág.47
10. Referencias	pág.48
11.1. Bibliografía	pág.48
11.2. Webliografía	pág.49
11. Anexos	pág.51

Índice de elementos

Gráficos

Gráfico 1. Comparación del tamaño de las diferentes bases de datos.....	pág.16
Gráfico 2. Análisis de sentimiento con <i>Azure</i> de <i>#NBAPlayoffs</i>	pág.31
Gráfico 3. Análisis de sentimiento con <i>Azure</i> de <i>#TiempoDePlayoffs</i>	pág.32
Gráfico 4. Análisis de sentimiento con <i>Azure</i> de <i>#NBAFinals</i>	pág.33
Gráfico 5. Análisis de sentimiento con <i>Azure</i> de <i>#LasFinalesDeTodaLaVida</i>	pág.34
Gráfico 6. Resultados análisis de sentimiento manual de <i>#LasFinalesDeTodaLaVida</i>	pág.36

Tablas

Tabla 1. Análisis de sentimiento con <i>Azure</i> de <i>#NBAPlayoffs</i>	pág.31
Tabla 2. Análisis de sentimiento con <i>Azure</i> de <i>#TiempoDePlayoffs</i>	pág.32
Tabla 3. Análisis de sentimiento con <i>Azure</i> de <i>#NBAFinals</i>	pág.33
Tabla 4. Análisis de sentimiento con <i>Azure</i> de <i>#LasFinalesDeTodaLaVida</i>	pág.34
Tabla 5. Análisis de sentimiento con <i>Azure</i>	pág.34
Tabla 6. Resultados análisis de sentimiento manual de <i>#LasFinalesDeTodaLaVida</i>	pág.36
Tabla 7. Análisis de sentimiento con <i>Azure</i> de <i>#LasFinalesDeTodaLaVida</i>	pág.37
Tabla 8. Resultados análisis de sentimiento manual de <i>#LasFinalesDeTodaLaVida</i>	pág.37
Tabla 9. Comparación de análisis manual con <i>Azure</i> en <i>#LasFinalesDeTodaLaVida</i>	pág.37

Tabla 10. Ejemplo de valoración de tweets entre análisis manual y automático.....	pág.39
Tabla 11. Resultados por clusters de XLStat de <i>#NBAPlayoffs</i>	pág.39
Tabla 12. Resultado de los clusters centroides de XLSTAT de <i>#NBAPlayoffs</i>	pág.40
Tabla 13. Resultados por clusters de XLStat de <i>#TiempoDePlayoffs</i>	pág.41
Tabla 14. Resultado de los clusters centroides de XLSTAT de <i>#TiempoDePlayoffs</i>	pág.41
Tabla 15. Resultados por clusters de XLStat de <i>#NBAFinals</i>	pág.43
Tabla 16. Resultado de los clusters centroides de XLSTAT de <i>#NBAFinals</i>	pág.43
Tabla 17. Resultados por clusters de XLStat de <i>#LasFinalesDeTodaLaVida</i>	pág.44
Tabla 18. Resultado de los clusters centroides de XLSTAT de <i>#LasFinalesDeTodaLaVida</i>	pág.44

Imágenes

Imagen 1. Resultado final de los enfrentamientos de los <i>Play-offs</i> 2022 de la NBA.....	pág.10
Imagen 2. Interfaz de <i>Twitter Archiver</i>	pág.14
Imagen 3. Base de datos del 7 de junio a las 12:00 conseguida con <i>Twitter Archiver</i>	pág.15
Imagen 4. Interfaz de <i>Azure</i>	pág.17
Imagen 5. Ejemplo de la puntuación obtenida por parte de <i>Azure</i> a algunos <i>tweets</i> de la base de datos de <i>#NBAFinals</i>	pág.17
Imagen 6. Extracto de los <i>tweets</i> que han sido seleccionados para el análisis de sentimiento manual.....	pág.18

Imagen 7. Interfaz <i>online</i> actual de <i>WordNet</i>.....	pág.20
Imagen 8. Procesamiento del Lenguaje Natural vs Lenguaje Natural.....	pág.21
Imagen 9. Página web de <i>Atribus</i>.....	pág.26
Imagen 10. <i>Topics</i> y <i>hashtags</i> extraídos de las conversaciones en <i>Social Media</i> sobre cápsulas de café internacionalmente.....	pág.27
Imágen 11. Índice de aceptación de materiales empleados en la fabricación de cápsulas de café.....	pág.28
Imagen 12. Resultados del análisis de sentimiento de la etiqueta <i>#Batlencia</i>.....	pág.29
Imagen 13. Mapa de conversación de los <i>clusters</i> de <i>#Batlencia</i>.....	pág.29
Imágen 14. Ejemplo de análisis demográfico de <i>Atribus</i>.....	pág.30

1. Introducción

Desde principios del siglo XXI las redes sociales han intentado llegar a nuestras vidas y para sus creadores ha sido un proceso costoso. Hasta que la sociedad no ha dado el salto a la modernización de sus teléfonos móviles y ha adoptado el *smartphone* como una forma de vida, estas no han podido formar parte de nuestro día a día. Desde entonces la sociedad ha pasado por redes sociales como *Messenger, MySpace, Tuenti, Fotolog, Snapchat, Musically...*

Tras todas estas plataformas, y muchas más, parece que la sociedad ha logrado estabilizarse en un pequeño grupo de redes sociales que hoy casi es un monopolio. En las manos del magnate estadounidense *Marck Zuckerberg* se encuentran las plataformas de *Facebook*, de la que es el creador, y *Whatsapp* e *Instagram*, estas dos últimas fueron compradas por él tras su éxito. Estas tres redes sociales cuentan con millones de descargas y son usadas a diario, pero hay más redes sociales, que quizás no son tan populares pero sí muy usadas, estas por ejemplo pueden ser *Tik Tok*, que desde hace unos dos años no para de experimentar un gran crecimiento de popularidad, *Pinterest*, que es una red social con un nicho muy específico pero que es mundialmente usada, o *YouTube* plataforma de videos bajo demanda perteneciente a *Google*.

Junto con todas estas plataformas hay otra que convive con ellas, *Twitter*. Los creadores de esta red social, *Evan Williams, Jack Dorsey* y *Biz Stone*, *"imaginaron un espacio de conversación en el que las personas entraran para contar qué estaban haciendo en ese momento. (...) Los usuarios decidieron hablar de otras cosas, pensaron utilizarlo para intercambiar información y para comentar lo que estaba pasando tanto a su alrededor como en cualquier otro lugar del mundo, y son los usuarios los que cada día descubren un nuevo uso que no estaba previsto"* (López Sobejano, 2012).

Tal ha sido el uso que los usuarios le han dado a *Twitter* que hoy en día es una herramienta indispensable para el seguimiento de cualquier tipo de acontecimientos (culturales, deportivos, políticos...). *"El uso de esta red social para el seguimiento de la actualidad ha llegado a tanto que los periodistas valoran Twitter como fuente informativa de relativa fiabilidad, aunque el 75 por 100 reconoce que la*

considera como tal, aunque con algunas prevenciones, pues más del 72 por 100 de los informadores que tienen esa consideración de la red social reconocen que contrasta la información que reciben a través de este canal con otras fuentes informativas que les ofrecen credibilidad” (Martínez-Fesneda Osorio y Sánchez Rodríguez, 2022).

La NBA (*National Basketball Association*) es la liga estadounidense de baloncesto profesional que en la pasada temporada, 21-22, cumplió 75 años desde su creación. Esta competición cuenta con un total de 36 equipos, dividiendo su clasificación en dos conferencias, este y oeste, y durante su temporada regular se juegan un total de 82 partidos. Tras esta, se disputa una competición de eliminación, que en el mundo del deporte se conoce como ‘*Play-offs*’, entre los primeros ocho clasificados de cada conferencia. Es decir, se hacen un total de 8 emparejamientos, cuatro por conferencia (1º vs 8º / 2º vs 7º / 3º vs 6º / 4º vs 5º) y cada emparejamiento juega al mejor de siete partidos, es decir, se clasifica a la siguiente ronda el primer conjunto que consiga cuatro victorias. Conforme avanzan las eliminatorias se proclamará un campeón de la conferencia este, otro de la oeste y ambos se enfrentarán en una serie a siete partidos para lograr ser campeón de la NBA. En la campaña 21-22 los *Playoffs* comenzaron el 17 abril y finalizaron el 17 de junio, proclamándose campeón el conjunto de los Golden State Warriors.

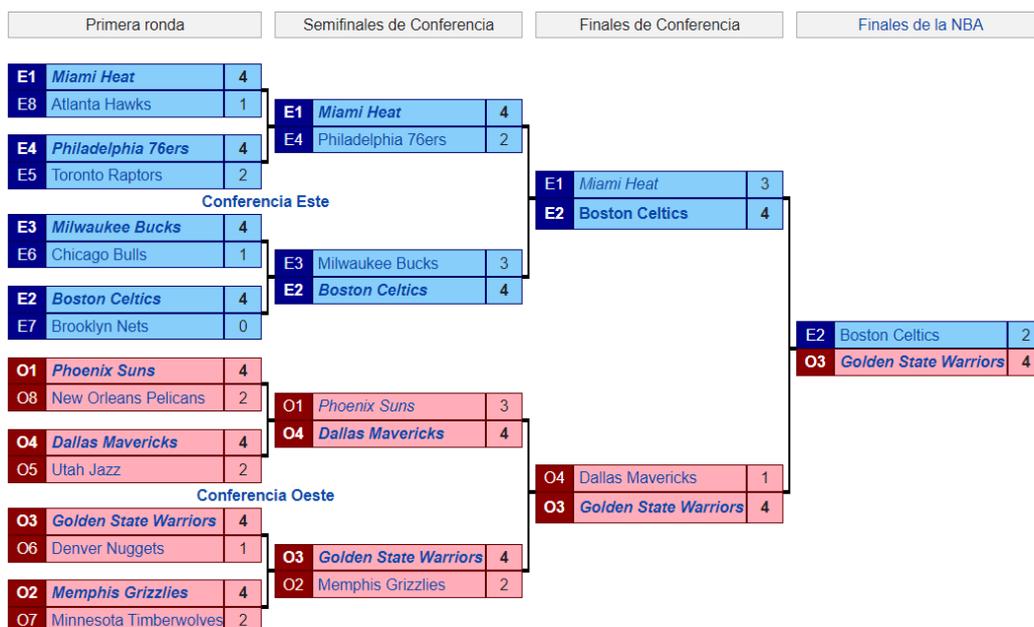


Imagen 1. Resultado final de los enfrentamientos de los *Play-offs* 2022 de la NBA / Fuente: Wikipedia.

Con estos elementos *Twitter* nos ofrece la posibilidad de poder analizar los *tweets* que los diferentes usuarios postean en esta red social, incluso también podemos analizar a los propios usuarios. Esto lo podemos hacer gracias a las siguientes dos metodologías de estudio: el análisis de sentimiento y el análisis de *clusters*.

El análisis de sentimiento es una metodología basada en la interpretación de textos con la ayuda de aplicaciones destinadas a este fin, las cuales están formadas por innumerables diccionarios. De estos, las aplicaciones descifran la intención de un mensaje en referencia al tema tratado en el texto. Esta metodología aplicación nos ayudará a conocer la opinión general de los usuarios de *Twitter* en referencia a un tema concreto.

El análisis de *clusters* en *Twitter* es una metodología la cual se realiza con el objetivo de conocer las características de los diferentes perfiles que *twitteen* sobre una tema para después poder aunarlos en diferentes grupos, *clusters*, y conseguir reducir los diferentes perfiles de usuarios que hablan sobre un tema en esta red social.

2. Motivación

Por lo estudiado en la asignatura de Audiencia Participativa y Social, en el Máster en Comunicación Transmedia, tanto mis compañeros como yo aprendimos las metodologías de análisis de sentimiento y de análisis de *clusters*, ambos con *softwares* destinados a estas tareas. Este nuevo aprendizaje me llamó mucho la atención y me parece un avance muy interesante a la hora de conocer la opinión de la sociedad.

La primera opción para este trabajo final de máster fue realizar un análisis de sentimiento de tweets en relación a la crisis interna del Partido Popular. En el momento en el que se seleccionó este tema se encontraron diferentes obstáculos, como la obtención de la API por parte de la plataforma Twitter para la extracción de los tweets necesarios.

A pesar de este inconveniente para cubrir el tema entero (de principio a fin de la crisis) se decidió abarcar solamente el final, la dimisión de Alberto Nuñez Feijoó,

como presidente de la Xunta de Galicia para ocupar su cargo como secretario general del Partido Popular.

Con este planteamiento se planificó una estrategia de extracción de tweets con la herramienta *Twitter Archiver*. El día de la dimisión y a la hora del acto del anuncio se comenzó la extracción de tweets cada cinco minutos durante una hora y diez minutos. Con esto se consiguieron extraer aproximadamente unos 1.100 tweets.

Una vez con el bruto de la base de datos se procedió a limpiarla y en este proceso ocurrieron dos problemas:

- Dentro de la búsqueda se colaron tweets acerca de la dimisión del presidente de la Federación Española de Rugby, quién también se apellida Feijóo.
- Entre las diferentes búsquedas se repitieron muchos tweets.

Con esta problemática se quedó una base de datos por debajo de los 200 tweets, lo que es insuficiente para la realización de un TFM. En este punto sí que tuvimos que cambiar por completo el tema del trabajo.

En este momento es cuando se seleccionó el tema actual, para así poder seguir trabajando con las metodologías elegidas.

3. Estructura

La estructura que vamos a seguir para este trabajo final de máster va a ser la siguiente:

- **Objetivos.** En este apartado hablaremos tanto de los objetivos principales como de los secundarios.
- **Metodología.** En este trabajo aplicaremos dos metodologías. Primero, el **análisis de sentimientos**, de este realizaremos el automático y el manual para así poder comparar resultados. Y, además, haremos un **análisis de clusters**, para conocer los tipos de usuarios que *tuitean* sobre el tema abordado en este trabajo.
- **Estado de la cuestión.** En este apartado abordaremos en qué situación se encuentran ambas metodologías que vamos a aplicar.

- **Análisis de datos.** Una vez aplicadas las metodologías elegidas, estas nos mostrarán unos datos los cuales tendremos que analizar.
- **Resultados.** Tras el análisis de los datos obtendremos unos resultados, los cuales explicaremos de forma clara y sencilla.
- **Conclusiones.** Una vez expuestos los resultados obtenidos tendremos que ser capaces de extraer unas conclusiones sólidas.
- **Relación del trabajo desarrollado con los estudios cursados.** Ya finalizada la tarea de investigación y habiendo mostrado sus resultados, explicaremos la relación que tiene con los estudios cursados.
- **Referencias.** El trabajo debe de estar apoyado en estudios anteriores los cuales estén avalados.
- **Anexos.** Para la realización de este trabajo se han necesitado cuatro bases de datos las cuales podrán ser consultadas en el enlace incluido en este apartado.

4. Objetivos

Los objetivos que tenemos en este trabajo los podemos dividir en principales y secundarios. Los objetivos principales son:

- Detectar posibles fallos y mejoras en la metodología de análisis de sentimiento la cual es considerada como 'reciente'.
- Conocer si las aplicaciones seleccionadas para este trabajo son las adecuadas para este tipo de análisis, puesto que las metodologías seleccionadas son muy nuevas en el campo de la investigación.

El objetivo secundario es:

- Conocer la opinión generalizada y el comportamiento en Twitter de los fans de la NBA acerca de las Finales de Conferencia y de las Finales de la NBA

5. Metodología

5.1. Generar la base datos

Tras seleccionar el actual tema, por las fechas en las que nos encontrábamos, vimos oportuno centrar el estudio del tema en las dos fases finales de los *Playoffs* de la NBA, es decir, en ‘Las Finales de Conferencia’ y en ‘Las Finales’.

Ya ubicados temporalmente decidimos hacer la selección de los *hashtags* que íbamos a utilizar para realizar la búsqueda de los *tweets* con la extensión de *Google Sheets, Tweet Archiver*. Esta extensión la utilizamos en la asignatura de Audiencia Participativa y Social, su función es poder descargar, mediante determinados parámetros, un archivo de *tweets* y verlos como una base de datos completamente manejable. Concretamente las búsquedas se pueden hacer por *hashtags* o por palabras y pueden filtrarse por idiomas.

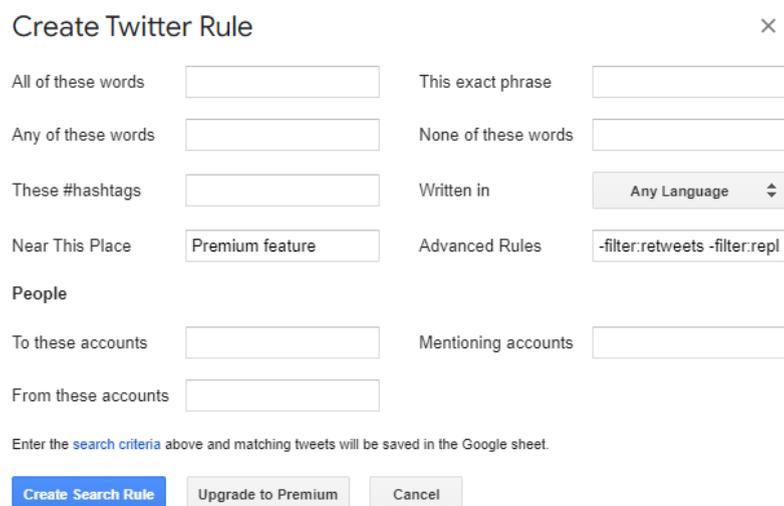


Imagen 2. Interfaz de *Twitter Archiver*. / Fuente: *Twitter Archiver*.

Una vez conocida la aplicación que íbamos a usar para obtener nuestra base de datos decidimos que lo mejor era hacer la búsqueda a través de *hashtags*, y no desde uno único, sino desde varios a la vez. Para empezar decidimos usar las etiquetas ‘*#TiempoDePlayoffs*’ y ‘*#NBAPlayoffs*’. Estos *hashtags* pertenecen, respectivamente, a la retransmisión española de Movistar+ y la propia NBA. Para cuando comenzasen ‘Las Finales’, a las dos anteriores se sumarían las etiquetas

'#LasFinalesDeTodaLaVida' y '#NBAFinals', también correspondientes a la cadena de Movistar+ y a la NBA.

Una vez seleccionada la aplicación que íbamos a usar para generar nuestra base de datos y los *hashtags* que íbamos a analizar decidimos establecer una patrón de búsquedas. Así, diariamente desde el 18 de mayo hasta el 17 de junio establecí tres búsquedas al día, una a las 12:00 horas, otra a las 17:00 horas y, una última a las 22:00 horas.

Twitter Query: #TiempoDePlayoffs lang:es -filter:retweets -filter:replies			
Date	Screen Name	Full Name	Tweet Text
05/06/2022 22:17	@pplu_nosgar	Pepelu Nosgar	Para ganar a Curry, hay que hacer demasiadas cosas bien, y es
05/06/2022 17:35	@yomismounomas1	yomismounomas	#TiempoDePlayoffs #LasFinalesDeTodaLaVida Ahora mismo B
04/06/2022 10:38	@MovistarNBA	NBA en Movistar Plus+	¿Qué diferencias hay entre un partido de liga regular y uno de
02/06/2022 20:23	@davidblanco_21	▲	ME VOY A COMPRAR UNA TABLET PARA PONER ESTE CUARTO E
02/06/2022 20:03	@cinefiliault	cinefiliault	Gracias Jaylen Brown. ¡Qué jugador! 🍌🍌🍌 #TiempodePlayo
02/06/2022 19:15	@NBAspain	NBA Spain	Los Splash Brothers ya son los dos jugadores con más triples e
02/06/2022 19:12	@cinefiliault	cinefiliault	Por primera vez en estos Playoffs NBA, veo problemas reales e
02/06/2022 18:51	@DavidHM_F1	DavidHM_F1	Se nos viene la locuuura con estas. #LasFinalesDeTodaLaVida i
02/06/2022 18:36	@javiiregalado	Javier Regalado	Es verdad que con tanto alto (Robert Williams, Horford, Theis)
02/06/2022 18:05	@pplu_nosgar	Pepelu Nosgar	De los @celtics desde shiquetito #tiempodeplayoffs
02/06/2022 17:45	@samy_rondo9	DURANTULA.	Vamosss Celticss caraaajoo!!! 🍀🍀🍀 #FinalesNBA #tiempodep
02/06/2022 07:21	@elsebikas	seb dice'	El verdadero TIC TAC 🗡️ #LasFinalesDeTodaLaVida #tiempode
31/05/2022 06:12	@MovistarNBA	NBA en Movistar Plus+	¡Cómo nos gusta el Flipómetro de #BasketAlDía! Las mejores j
30/05/2022 15:38	@carlosmaxwell	Carlos Maxwell	¡Felicidades @alhorford! Primer dominicano que jugará en una
30/05/2022 15:37	@lionoga	lionoga	#TiempoDePlayoffs lo de "el día de la mancuerna en Miami" h
30/05/2022 07:09	@Alfred_Castilla	Alfredo_Castilla	4-0 a Brooklyn Nets en primera ronda 4-3 a los campeones Bu
30/05/2022 07:07	@Alfred_Castilla	Alfredo_Castilla	JAYSON TATUM Vs STEPHEN CURRY. Boston Celtics vs Golden S
30/05/2022 05:40	@ElMundoNBA	NBA	🔥 Final de la NBA 🔥 🇺🇸 Tu a Boston y yo a California 🇺🇸 #NBA
30/05/2022 03:55	@basket_europe	BasketEuropeWorld	🟢🟢 Ya disponible nuevo contenido en nuestra carpeta del s
30/05/2022 02:46	@samuelljackson	samuel L jacksson	Bueno, tenemos otra cita para las proximas semanas @rodrigi
30/05/2022 01:12	@Alfred_Castilla	Alfredo_Castilla	Serán unas finales muy igualadas, Boston con la mejor defens

Imagen 3. Base de datos del 7 de junio a las 12h conseguida con *Twitter Archiver* / Fuente: Elaboración propia.

Como se ha explicado anteriormente, no se empezaron las búsquedas con todos los *hashtags*, así que estas son las fechas en las que cada etiqueta fue sumándose al proceso de búsqueda:

-18 de mayo: '#TiempoDePlayoffs' y 'NBAPlayoffs'

-27 de mayo: 'FinalsNBA'

-3 de junio: '#LasFinalesDeTodaLaVida'

Una vez finalizaron los Playoffs de la NBA, el 17 de junio, y ya habíamos generado la base de datos, el paso siguiente era limpiarla, pues detectamos que la aplicación *Twitter Archive* tiene el defecto de que descarga muchos *tweets*

duplicados. Esta limpieza se hizo con la herramienta ‘Borrar duplicados’ de *Microsoft Excel*. Tras este proceso nuestra base de datos ya estaba completamente lista para ser sometida a los dos análisis que planteamos en este estudio.

Puntualizar que nuestra base de datos está formada por un total de 9.862 *tweets* y, concretamente la etiqueta ‘*#TiempoDePlayoffs*’ tiene 1.107 *tweets*, el *hashtag* ‘*#NBAPlayoffs*’ tiene un total de 3.374, ‘*#NBAFinals*’ tiene 4.680 *tweets* y, por último, el *hashtag* ‘*#LasFinalesDeTodaLaVida*’ tiene un total de 732 *tweets*.

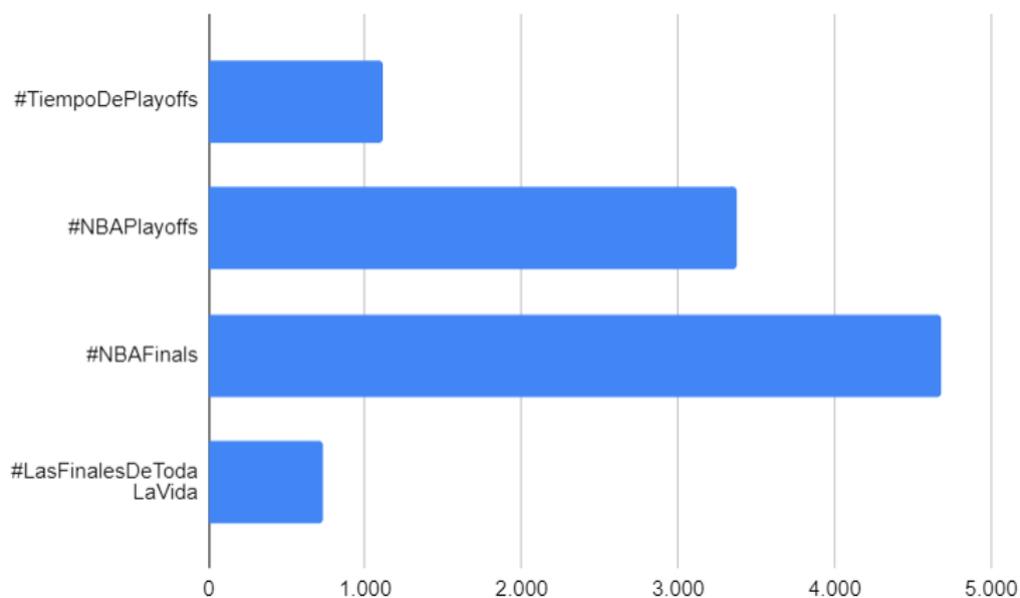


Gráfico 1. Comparación del tamaño de las diferentes bases de datos. / Fuente: Elaboración propia.

5.2. Análisis de sentimiento automático

Una vez tenemos generada nuestra base de datos el paso siguiente era encontrar una aplicación de análisis de sentimiento a la que pudiésemos adjuntar nuestros datos para analizarlos. Tras un proceso de investigación y de diferentes aplicaciones, seleccionamos el complemento ‘*Azure*’ de *Microsoft*, el cual tiene una función de ‘*Text Sentiment Analysis*’, que la usamos para realizar de forma automática el análisis de sentimiento.

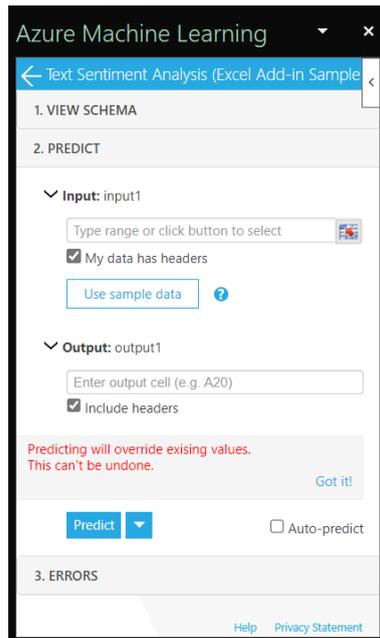


Imagen 4. Interfaz de Azure / Fuente: Elaboración propia.

'Azure' es un servicio basado en el procesamiento del lenguaje natural (PNL) para comprender y analizar textos. Es capaz de realizar sus funciones en hasta 115 idiomas, para ello 'Azure' se hace servir de una función de detección de idiomas, el cual está formado por una serie de algoritmos de aprendizaje y de inteligencia artificial. La última vez que este algoritmo fue actualizado fue el 05/01/2021.

Dentro de sus diferentes funciones de análisis de texto se encuentra el análisis de sentimiento, que lo realiza de la siguiente forma. Cuando analiza los textos asigna a cada uno un valor de entre 0 y 1, siendo los valores más próximos al 0 los clasificados como 'negativos', los más cercanos al 1 como 'positivos' y los que se encuentran cerca del 0,5 son etiquetados como 'neutros'.

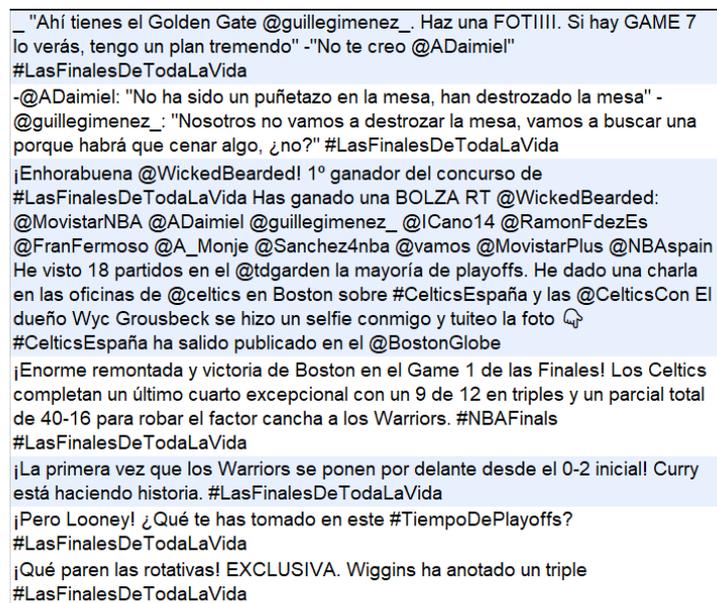
A por el Titulo, Dub Nation. 🏆 #WarriorsVsMavs #NBAFinals #NBAPlayoffs RT @warriors: DUB NATION STAND UP The Golden State Warriors are Western Conference Champions	positive	0.804336726665497
Lo de los Warriors no tiene otra explicación. Son el mejor equipo del mundo. Los continuos carretones, las puertas atrás, la facilidad de Draymond Green para ver el espacio libre de sus compañeros... Es una gozada verles jugar. Un equipo imparabile. #NBAFinals #DubNation	neutral	0.56373792886734
Como no alegrarse por este man, vuelve a su lugar, las #NBAFinals RT @NBA: The FIRST EVER Earvin "Magic" Johnson Trophy recipient and the Western Conference Finals M-V-P... @StephenCurry30 🔥 Check out some of his best plays from the Western Conference Finals!	positive	0.922272801399231
Noviembre era y la de Poole no la vimos venir nadie. Se vienen unas #NBAFinals de la hostia. RT @MonchoFdezP: Sigilosos, pero con firmeza e ideas claras. Mucho cuidado a los Warriors 21/22, más aún con la vuelta de Klay.	positive	0.809069275856018
✓ Golden State Warriors vence en el quinto encuentro a Dallas Mavericks y vuelven a las 🏆 #NBAFinals 3 años después. 🏆 GSW 120 - 110 DALLAS MAVERICKS 🏆 ¿Conseguirán el séptimo anillo de su historia? [Via @NBAspain]	positive	0.704888999462128

Imagen 5. Ejemplo de la puntuación obtenida por parte de Azure a algunos tweets de la base de datos de #NBAFinals / Fuente: Elaboración propia.

5.3. Análisis de sentimiento manual

El análisis de sentimiento automático es una metodología considerada como nueva y uno de los objetivos de este trabajo es encontrar posibles fallos y mejoras en ella. Por esto, vemos necesario enfrentar los resultados del análisis automático con los de un resultado manual.

Para ello vamos a coger los *tweets* del *hashtag* '#LasFinalesDeTodaLaVida', ya que se trata de la etiqueta con menos *tweets* y haremos la clasificación de estos en positivo, negativo o neutro. A partir de estos resultados haremos una comparación de ambos resultados para conocer el margen de fiabilidad que las herramientas de análisis de sentimiento tienen a día de hoy.



"Ahí tienes el Golden Gate @guillegimenez. Haz una FOTIII. Si hay GAME 7 lo verás, tengo un plan tremendo" -"No te creo @ADaimiel"
#LasFinalesDeTodaLaVida
-@ADaimiel: "No ha sido un puñetazo en la mesa, han destrozado la mesa" -
@guillegimenez_: "Nosotros no vamos a destrozarnos la mesa, vamos a buscar una porque habrá que cenar algo, ¿no?" #LasFinalesDeTodaLaVida
¡Enhorabuena @WickedBearded! 1º ganador del concurso de
#LasFinalesDeTodaLaVida Has ganado una BOLZA RT @WickedBearded:
@MovistarNBA @ADaimiel @guillegimenez_ @ICano14 @RamonFdezEs
@FranFermoso @A_Monje @Sanchez4nba @vamos @MovistarPlus @NBAspain
He visto 18 partidos en el @tdgarden la mayoría de playoffs. He dado una charla en las oficinas de @celtics en Boston sobre #CelticsEspaña y las @CelticsCon El dueño Wyc Grousbeck se hizo un selfie conmigo y tuiteo la foto 📷
#CelticsEspaña ha salido publicado en el @BostonGlobe
¡Enorme remontada y victoria de Boston en el Game 1 de las Finales! Los Celtics completan un último cuarto excepcional con un 9 de 12 en triples y un parcial total de 40-16 para robar el factor cancha a los Warriors. #NBAFinals
#LasFinalesDeTodaLaVida
¡La primera vez que los Warriors se ponen por delante desde el 0-2 inicial! Curry está haciendo historia. #LasFinalesDeTodaLaVida
¡Pero Looney! ¿Qué te has tomado en este #TiempoDePlayoffs?
#LasFinalesDeTodaLaVida
¡Qué paren las rotativas! EXCLUSIVA. Wiggins ha anotado un triple
#LasFinalesDeTodaLaVida

Imagen 6. Extracto de los *tweets* que han sido seleccionados para el análisis de sentimiento manual. / Fuente: Elaboración propia.

5.4. Análisis de clusters

Para realizar el análisis de *clusters* hemos cogido una copia de las bases de datos que ya habíamos realizado previamente. Hemos seleccionado la aplicación 'XLStat' (software de análisis avanzado de datos en Excel) puesto que es una aplicación que hemos usado en el máster y ya sabemos utilizar, por lo cual, también sabemos que es un *software* que va a solventar nuestras necesidades.

Una vez hemos activado este *software* dentro de nuestro Excel tendremos que seleccionar la opción 'Análisis de datos' y dentro de este seleccionamos

'Clusters de K-means'. Se nos abrirá una ventana emergente y en el cuadro de diálogo 'Tabla de observaciones/variables' lo cumplimentaremos seleccionando las columnas de 'seguidores', 'seguidos', 'retweets' y 'favoritos'. Por defecto, esta ventana emergente viene con las opciones 'Etiquetas de las columnas' y 'Hoja' seleccionadas, las tendremos que dejar así. Y también vienen cumplimentadas por defecto los cuadros de diálogo de 'Índice de disimilitud' y de 'Criterio de clasificación', los cuales los dejaremos como están.

Una vez ya hayamos seguido estos pasos hacemos *click* en 'Ok', paso que repetiremos cuando nos aparezca la siguiente ventana emergente. Una vez realizados estos dos pasos, *Excel* nos generará una nueva hoja con el nombre '*K-means1*' donde podremos ver los resultados de análisis de la aplicación '*XLStat*'.

Este paso lo haremos con las cuatro bases de datos de los cuatro *hashtags* y dentro de los resultados que nos arroja buscaremos dos tablas, las cuales están nombradas como:

- *Results by clusters*
- *Clusters centroids*

De estas dos tablas extraeremos los datos que tendremos que analizar.

6. Estado de la cuestión

Una vez expuestas las metodologías que aplicaremos en este trabajo de investigación, estas deben ser justificadas con su debido marco teórico. Primero procederemos a la justificación del análisis de sentimiento y, después, el análisis de *clusters*.

6.1. Análisis de sentimiento

¿Cómo surge?

Es importante conocer cómo surge esta metodología, la primera referencia que hemos encontrado en cuanto al análisis de textos mediante el uso de un *software* es de 1998, en un artículo desarrollado en la Universidad de Ottawa (Scott, S., & Matwin, S.,1998). En este trabajo de investigación usaron el programa '*WordNet*', el cual tenía un gran *background* en conocimientos léxicos básicos.

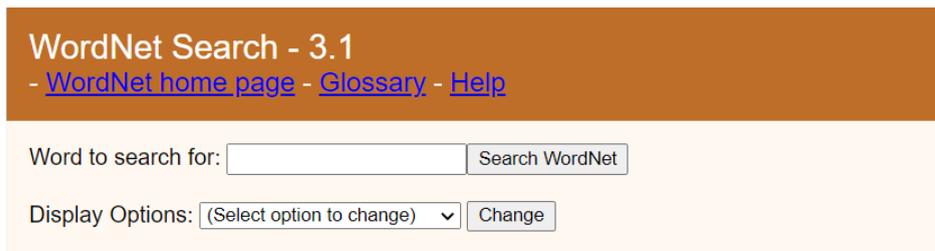


Imagen 7. Interfaz *online* actual de *WordNet*. / Fuente: *WordNet*.

Realmente este trabajo no fue usado para analizar opiniones generales sobre un tema, sino para intentar solucionar los problemas de léxico que otras aplicaciones de análisis de sentimiento tenían.

A día de hoy la aplicación usada en este trabajo de 1998 sigue en funcionamiento y se trata de una base de datos de léxico, pero únicamente en inglés.

¿Cómo lo conocemos a día de hoy?

El análisis de sentimiento es una metodología la cual nos permite conocer de forma generalizada la opinión de los usuarios de la red, en este caso los de *Twitter*. *“Su objetivo es reconocer y evaluar el valor emocional existente detrás de los textos analizados, a través de su estructura, clasificándolos en positivos, negativos o neutro. En la actualidad, esta metodología se aplica principalmente en la interpretación de los textos difundidos en medios sociales como Twitter”*, (Arcila-Calderón, C., Ortega-Mohedano, F., Jiménez-Amores, J., y Trullenque. (2017). Así nos lo define este estudio que aplicó la misma metodología para su trabajo de investigación, en este caso centrado en el ámbito de la política.

“Generalmente el análisis de sentimiento atiende a dos cometidos, en primer lugar, reconocer las expresiones de sentimiento y definir la orientación del sentimiento expresado por los individuos”, (C. Honey and S.C. Herring, 2009). Esta definición también apoya la intencionalidad que tenemos en este trabajo al aplicar la metodología del análisis de sentimiento.

Para conseguir la evaluación de estos mensajes de twitter existen diversas técnicas, *“las principales son las basadas en el procesamiento del lenguaje natural (PLN) y las que se basan en diccionarios”* (Medhat, W., Hassan, A. & Korashy, H., 2014).

¿Qué es el Procesamiento del Lenguaje Natural (PLN)?

Antes de conocer qué es el Procesamiento del Lenguaje Natural, primero debemos conocer qué es lo que la comunidad científica determina como 'lenguaje natural'. *“El lenguaje natural (LN) es el medio que utilizamos de manera cotidiana para establecer nuestra comunicación con las demás personas”* (Vásquez, A. C., Quispe, J. P., & Huayna, A. M., 2009).

Una vez conocemos la acepción de 'lenguaje natural' aceptada por la comunidad científica, ya podemos pasar a la definición de 'Procesamiento del Lenguaje Natural'. *“El PLN consiste en la utilización de un lenguaje natural para comunicarnos con la computadora, debiendo ésta entender las oraciones que le sean proporcionadas, el uso de estos lenguajes naturales, facilita el desarrollo de programas que realicen tareas relacionadas con el lenguaje o bien, desarrollar modelos que ayuden a comprender los mecanismos humanos relacionados con el lenguaje”* (Vásquez, A. C., Quispe, J. P., & Huayna, A. M., 2009).

Por lo tanto, y tras estas dos definiciones entendemos que el lenguaje natural es el que la sociedad utiliza en su día a día y que el procesamiento del mismo es necesario para que la comunicación entre ordenadores y humanos sea posible.

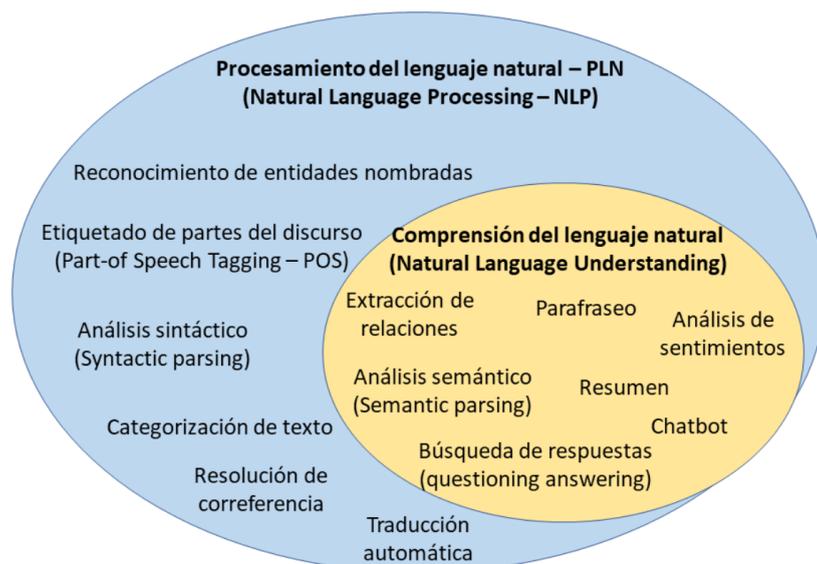


Imagen 8. Procesamiento del Lenguaje Natural vs Lenguaje Natural. / Fuente: Wikipedia.

¿Qué es el análisis de sentimiento basado en diccionarios?

Este tipo de análisis de sentimiento consiste en “*un listado de términos (pueden ser palabras o multipalabras) que tienen asociados una determinada orientación de sentimiento. De acuerdo con la distinción inicial sobre el tipo de análisis posible, esta orientación puede ser de polaridad (positiva / negativa), de intensidad de sentimiento o de tipo de emoción. El algoritmo que evalúa la orientación de sentimiento únicamente tiene en cuenta la presencia en el texto del tweet, de términos incluidos en el diccionario*” (Baviera, T. 2017).

Según esta definición entendemos que los *software* destinados al análisis de sentimiento de textos mediante un *background* de sintaxis formado por diccionarios simplemente valoran las palabras positivas y negativas que se encuentran dentro de los textos para finalmente valorarlos.

¿Qué aplicaciones disponemos a día de hoy para el análisis de sentimiento?

Puesto que ya hemos hecho una diferenciación de los tipos de análisis de sentimiento vamos a conocer algunos de los programas destinados para este fin pero separándolos según la metodología que empleen.

Softwares basados en el PLN

-QuestionPro

QuestionPro es un *software* destinado a empresas para poner a su disposición herramientas de comunicación, tanto internas como externas. Para esto, ponen al servicio de sus clientes herramientas como las encuestas, las investigaciones de mercado o la apertura de canales de comunicación entre los diferentes niveles de responsabilidad de las empresas usuarias de *QuestionPro*.

Este *software* usa el Procesamiento Natural del Lenguaje a la hora de evaluar las respuestas de las encuestas ya que estas cuentan con la opción de ser de respuesta abierta para llegar más en profundidad a la opinión del consumidor.

-Google Cloud Natural Language

Google tiene su propio servicio de análisis de textos y para ello ofrece hasta tres productos diferentes. Estos son '*AutoML*', '*API de Natural Language*' y '*API de Natural Language de Healthcare*'.

El primero, '*AutoML*' es un *software* que mediante la inteligencia artificial (IA) puede entrenar al algoritmo a gusto del cliente para después usarlo como herramienta de análisis de sentimiento.

El segundo, '*API de Natural Language*', está basado en modelos previamente entrenados para el desarrollo de las tareas de análisis de sentimiento, opinión, clasificación de contenido, etc.

Por último, '*API de Natural Language de Healthcare*' es un producto diseñado para la extracción de información relevante de textos dedicados al ámbito de la medicina, sanidad...

-MeaningCloud for Excel

MeaningCloud es un complemento de *Microsoft Excel* que entre muchas de sus funciones una de ellas es el Análisis de textos. Dentro de esta ofrece herramientas para la clasificación de textos, el análisis de sentimiento, la identificación del idioma, la extracción de temas, la agrupación de textos y la categorización profunda.

Centrándonos en el tema que nos ocupa, el análisis de sentimiento, este complemento cuenta con una API de Análisis de Sentimiento propio. Esto permite que la aplicación extraiga sentimientos basados en aspectos, distinga hechos y opiniones, detecte ironía y desacuerdo. Con ella es posible también que los usuarios puedan definir sus propios diccionarios y su propio modelo de opinión.

-Shaip

Shaip es una empresa dedicada al entrenamiento de inteligencia artificial (IA). En estos entrenamiento de IA desarrollaron una API basada en procesamiento del lenguaje natural (PLN) y gracias a ella ofrecen servicios de análisis de sentimiento de marca.

Sus análisis llevan a cabo las mediciones de distintos tipos de parámetros, concretamente, polaridad, emociones, urgencia e intención. Esta empresa llega a asegurar hasta un 95% de fiabilidad en sus resultados.

Softwares basados en diccionarios

-SentiStrength

Esta aplicación de análisis de sentimiento basada en diccionarios es capaz de evaluar los textos desde bases de datos externas al software. Adjunto al mismo existe un documento en formato *'txt'* en el cual se incluyen palabras que van asociadas a una valoración para que el *software* sea capaz de valorar los textos. También existen archivos complementarios para poder entender y valorar el uso de emoticonos, la polaridad de las palabras, la intensidad del sentimiento y el uso de coloquialismos (jerga).

-WordNet-Affect

Esta aplicación, propiedad del dominio *WordNet* es capaz, mediante un *background* basado en diccionarios, de identificar emociones (ira, tristeza, sorpresa, etc.) y, por supuesto, de asignar valores a los textos, pero incluye una nueva valoración. El resto de aplicaciones que hemos visto hasta ahora calificaban los textos como positivos, neutros y negativos, pero *WordNet-Affect* incluye la ambigüedad como una valoración para sus análisis.

6.2. Clustering

¿Cómo surge y qué es el *clustering*?

El término *cluster* es en origen, un término económico, y el economista Michael E. Porter de la siguiente definición:

“Un clúster o cúmulo es un grupo geográficamente denso de empresas e instituciones conexas, pertenecientes a un campo concreto, unidas por rasgos comunes y complementarias entre sí.” (Michael E. Porter, 1999).

Cómo es obvio, la cuestión económica no es la que nos ocupa, por lo que el término *cluster* ha ido evolucionando. A día de hoy un *cluster* se entiende como “una

colección de objetos que son similares entre ellos y diferentes respecto a los miembros de otros grupos” (Gallardo Campos, M., 2009).

Para la colección de estos objetos, y por lo tanto para la formación de un *cluster*, es necesario el proceso que los conforma, que recibe el nombre de *clustering* y se entiende como “*el proceso de agrupar objetos, de tal modo que los objetos de un mismo grupo son más similares unos a otros que con objetos de otros grupos. (...) Es útil en muchas técnicas de aprendizaje automático.*” (Gallardo Campos, M., 2009).

Con esta definición de *clustering* nos surge el término de ‘técnicas de aprendizaje automático’ el cual podemos explicar como “*esquemas que se basan en el establecimiento de un modelo explícito o implícito que permita los patrones analizados para ser categorizados*”. (García-Teodoro, P., J. Díaz-Verdejo, G.Maciá-Fernández and E. Vázquez, 2009).

Con las definiciones reunidas podemos esclarecer que el *clustering* es un proceso basado en técnicas de aprendizaje automático el cual es capaz de reconocer características comunes en diferentes objetos para acabar formando diferentes grupos con similares atributos. Esta técnica se emplea a día de hoy en las redes sociales para encontrar usuarios (objetos) con patrones de comportamiento similares y así poder extraer datos que aunque pueden ser muy generales, también son muy concretos ya que se realiza una segmentación de los diferentes usuarios.

Atribus, Social Listening

Para el desarrollo de este trabajo hemos querido conocer como trabajan las agencias de *marketing* a la hora de emplear las metodologías de análisis de sentimiento y *clustering*. Con este objetivo nos hemos puesto en contacto con *Atribus*, la única plataforma que se dedica al *Social Listening* en España.

Social Listening

Plataforma de social intelligence

Descubre **cómo se comportan tus consumidores** en el ecosistema digital.

Extraemos datos de: 



Imagen 9. Página web de Atribus. / Fuente: Atribus.com.

Según *Atribus*, el *social listening* es “el uso de herramientas de monitorización de internet y redes sociales” para así poder encontrar diversos datos, relacionados con nuestra marca y nuestros clientes, para *a posteriori* poder crear estrategias de marketing *on-line* basadas en datos objetivos. Los datos que pueden ser requeridos para la creación de dichas estrategias pueden ser las menciones en redes sociales, la opiniones de nuestro público, la relevancia que las mismas hayan podido tener, etc.

Esta herramienta internacional, con dos sedes en España (Valencia y Madrid) ofrece servicios *on-line* como el análisis de campañas orgánicas, la gestión de la reputación *on-line* o la monitorización de consumidores, entre otras tareas. *Atribus* trabaja en cinco sectores muy específicos los cuales son la político, el turismo, la alimentación y el deporte.

Atribus ha trabajado con empresas de gran relevancia en dichos sectores, algunos ejemplos pueden ser el *Valencia CF* (deporte), *Just Eat* (alimentación) o *Segittur* (turismo). Gracias al contacto que hemos tenido con la plataforma hemos podido acceder de forma íntegra a dos informes de *social listening*. Concretamente vamos a ver y comentar un caso relacionado con el mundo de la alimentación y otro enmarcado en el ámbito deportivo.

Caso nº1. Nuevas formas de consumir café

Este informe de *social listening* sobre el consumo de café es de ámbito internacional, aunque los análisis llevados a cabo se han realizado únicamente sobre textos escritos en inglés y español. Se contó con más de 800.000 opiniones de un total de 277.382 usuarios. Para la realización de este estudio se emplean diversas metodologías, entre las cuales se encuentran el análisis de sentimiento y el *clustering*.

Es necesario aclarar que este estudio sobre el consumo del café busca conocer la opinión que existe en la sociedad sobre el consumo de café en formato cápsula.

Atribus realizó un análisis de *clusters* y detectó en el comportamiento de los usuarios el uso de determinadas palabras clave (*topics*) y el uso de *hashtags* relevantes en la conversación.

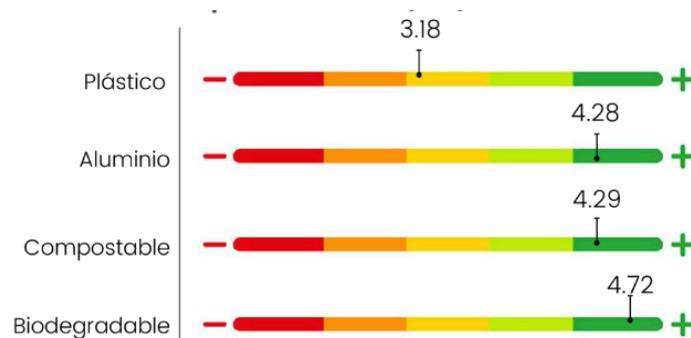
Topics	#Hashtags
<ul style="list-style-type: none">• Coffee pods• Coffee• Coffeepods• Pod• Cup coffee pods from keurig• Cup• Coffeelover• Nespresso• Best bean• Capsule	<ul style="list-style-type: none">#coffee#coffeepods#compostable#ad#coffeemachine#espresso#EspressoMachine#Nespresso#Sustainability#BostonBest

Imagen 10. Topics y hashtags extraídos de las conversaciones en Social Media sobre cápsulas de café internacionalmente. / Fuente: Atribus.

Después, en el informe podemos ver los resultados de diversos análisis de sentimiento referente a las cápsulas de café. Los análisis realizados fueron:

- Valoración de las cápsulas de café como formato de consumo (positivo, neutro o negativo).
- Valoración de los diferentes materiales que las componen (plástico, aluminio, compostable y reutilizable).

- Valoración de los diferentes atributos del café en cápsula (sabor, calidad, aroma, o acidez).



Imágen 11. Índice de aceptación de materiales empleados en la fabricación de cápsulas de café. / Fuente: Atribus.

Gracias a otras técnicas de *social listening* este informe incluye también, entre otros:

- Un estudio de mercado en el cual se realiza un índice de reputación entre dos marcas punteras del sector.
- Un análisis del consumidor en el que se habla de la demografía de estos (lugar de residencia, género y edad).
- Un estudio sobre los momentos de consumo del café.

Caso nº2. Valencia CF, HBOmax y The Batman

Este informe de *social listening* trata de una colaboración entre la entidad deportiva *Valencia CF*, la plataforma de *streaming HBOmax* y la película *The Batman*. De esta estrategia de *marketing* nace el *hashtag #Batlencia* y serán las conversaciones que incluyan esta etiqueta las que fueron analizadas por *Atribus*. Destacar que se trata de un estudio de menores dimensiones, ya que únicamente participaron 270 usuarios.

Al igual que en el estudio anterior, en este informe se han utilizado diferentes metodologías pertenecientes al *social listening* y entre ellas, como es obvio, se encuentran el análisis de sentimiento y el de *clusters*.

En relación al análisis de sentimiento podemos ver que *Atribus* tuvo en cuenta diferentes emociones que analizar (alegría, anticipación, confianza, disgusto,

Además de estos dos análisis, en el informe se incluye un análisis demográfico de los usuarios participantes en la conversación (edad, sexo y lugar de residencia), se muestran cuales han sido los usuarios más relevantes que han participado y cuales han sido los mensajes más destacados, es decir, con más repercusión e interacción.

Demographic profile in social media

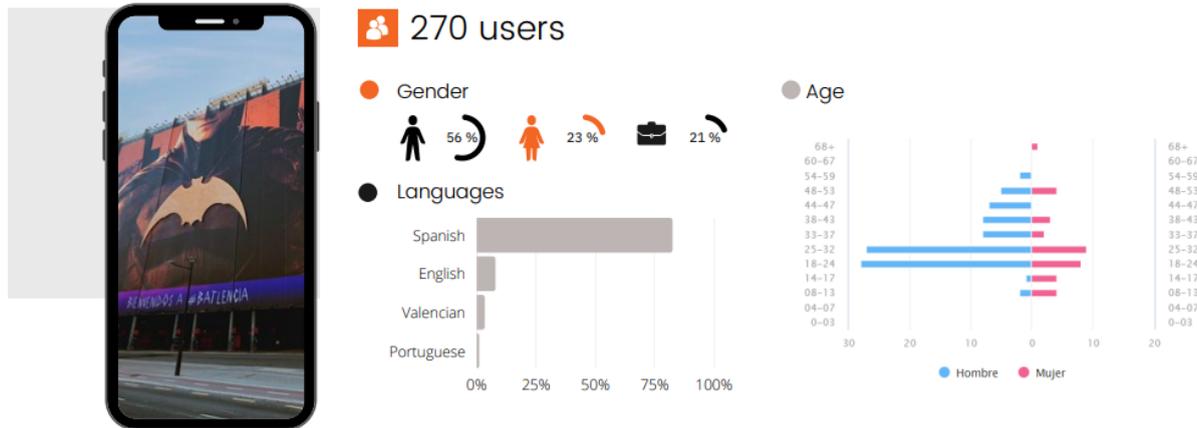


Imagen 14. Ejemplo de análisis demográfico de Atribus / Fuente: Atribus.

7. Resultados y discusión

Como hemos explicado en anteriores apartados, hemos generado un total de cuatro bases de datos, por lo que en este momento nos disponemos a explicar los resultados obtenidos después de haberlas sometido a los dos softwares seleccionados para este estudio.

7.1. Resultado del análisis de sentimiento automático

Después de haber analizado nuestra base de datos con la extensión de *Microsoft Excel*, *Azure*, los datos obtenidos en los diferentes *hashtags* son los siguientes.

En la etiqueta *#NBAPlayoffs* recopilamos un total de 3.374 *tweets* (después de limpiar la base de datos). *Azure* nos ha arrojado estos datos en el siguiente gráfico de barras:

Análisis de sentimiento #NBAPlayoffs

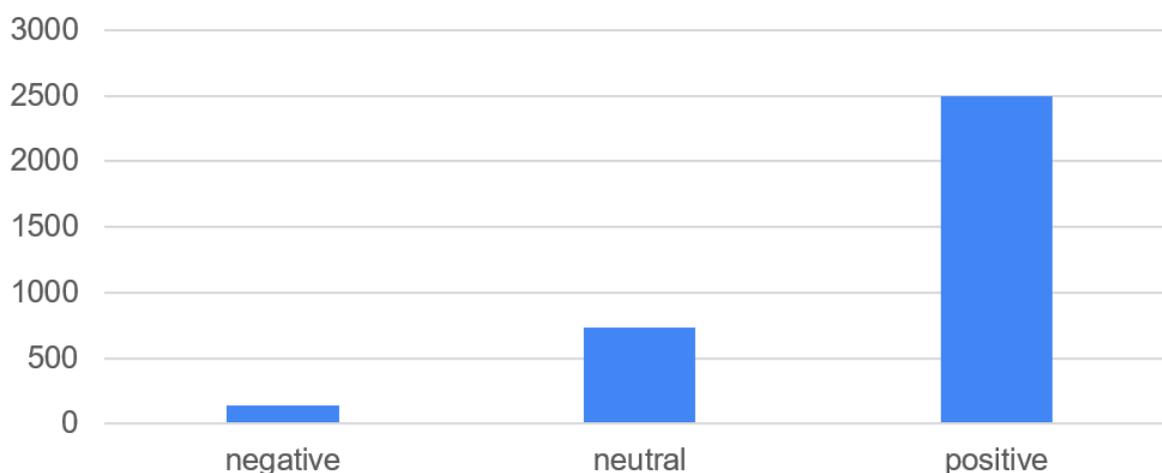


Gráfico 2. Análisis de sentimiento con Azure de #NBAPlayoffs / Fuente: Twitter / Azure / Elaboración propia.

Concretamente los datos expuestos en esta gráfica de barras son los siguientes:

Sentimiento	Total	Porcentaje
Negative	144	4,27%
Neutral	732	21,70%
Positive	2.497	74,01%
Total	3.374	100%

Tabla 1. Análisis de sentimiento con Azure de #NBAPlayoffs / Fuente: Twitter / Azure / Elaboración propia.

En la etiqueta *#TiempoDePlayoffs* recopilamos un total de 1.107 tweets (después de limpiar la base de datos). Azure nos ha arrojado estos datos en el siguiente gráfico de barras:

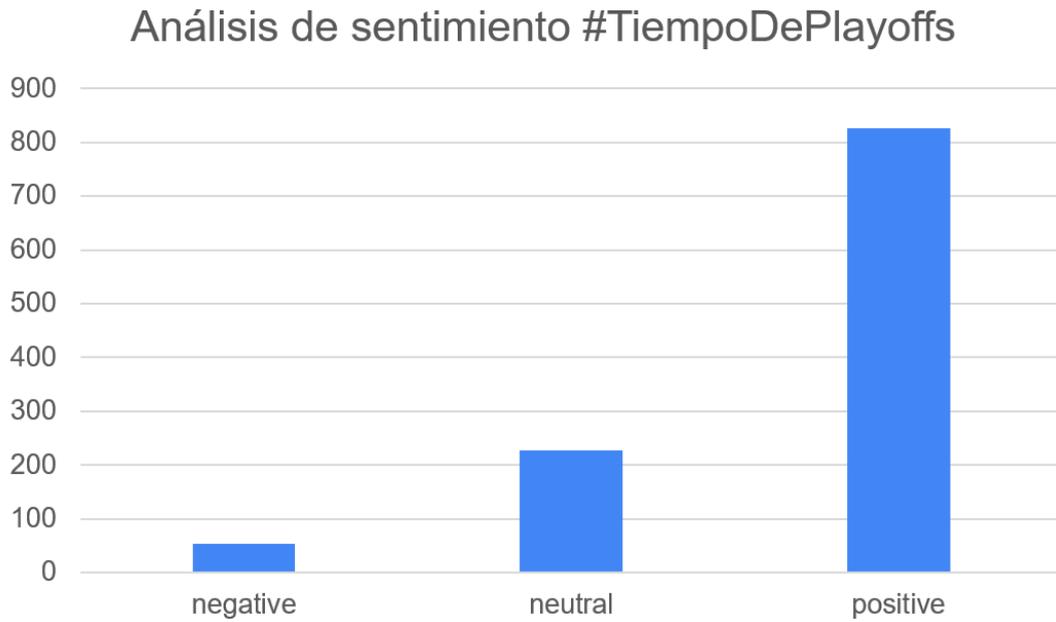


Gráfico 3. Análisis de sentimiento con Azure de #TiempoDePlayoffs / Fuente: Twitter / Azure / Elaboración propia.

Concretamente los datos expuestos en esta gráfica de barras son los siguientes:

Sentimiento	Total	Porcentaje
Negative	53	4,70%
Neutral	228	20,23%
Positive	826	73,29%
Total	1.107	100%

Tabla 2. Análisis de sentimiento con Azure de #TiempoDePlayoffs / Fuente: Twitter / Azure / Elaboración propia.

En la etiqueta #NBAFinals recopilamos un total de 4.680 tweets (después de limpiar la base de datos). Azure nos ha arrojado estos datos en el siguiente gráfico de barras:

Análisis de sentimiento #NBAFinals

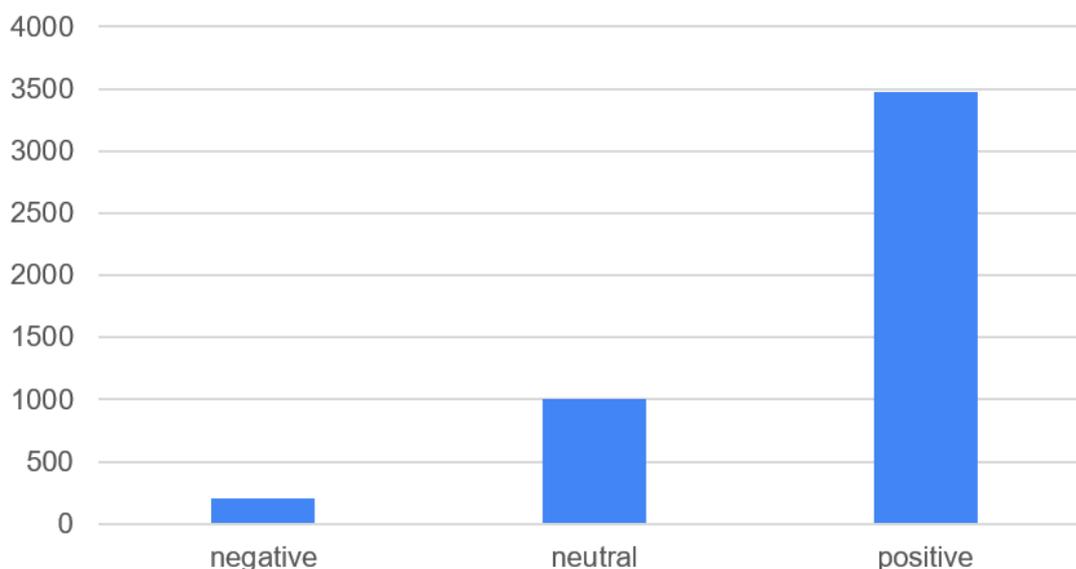


Gráfico 4. Análisis de sentimiento con Azure de #NBAFinals / Fuente: Twitter / Azure / Elaboración propia.

Concretamente los datos expuestos en esta gráfica de barras son los siguientes:

Sentimiento	Total	Porcentaje
Negative	203	4,34%
Neutral	1.002	21,41%
Positive	3.475	74,25%
Total	4.680	100%

Tabla 3. Análisis de sentimiento con Azure de #NBAFinals / Fuente: Twitter / Azure / Elaboración propia.

En la etiqueta *#LasFinalesDeTodaLaVida* recopilamos un total de 4.680 tweets (después de limpiar la base de datos). Azure nos ha arrojado estos datos en el siguiente gráfico de barras:

Análisis de sentimiento
#LasFinalesDeTodaLaVida

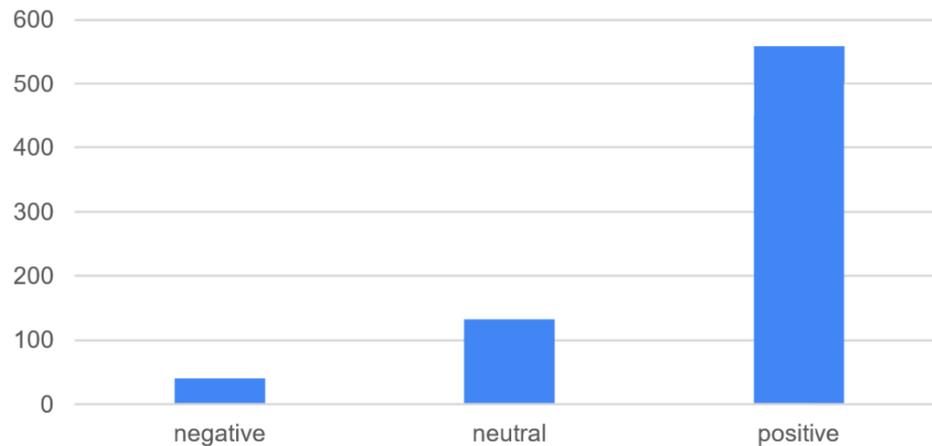


Gráfico 5. Análisis de sentimiento con Azure de #LasFinalesDeTodaLaVida / Fuente: Twitter / Azure / Elaboración propia.

Concretamente los datos expuestos en esta gráfica de barras son los siguientes:

Sentimiento	Total	Porcentaje
Negative	40	5,46%
Neutral	133	18,17%
Positive	559	76,36%
Total	732	100%

Tabla 4. Análisis de sentimiento con Azure de #LasFinalesDeTodaLaVida / Fuente: Twitter / Azure / Elaboración propia.

Una vez expuestos los datos repartidos por *hashtags* podemos hacer una visión completa de los mismos.

Sentimiento	Total	Porcentaje
Negative	440	4,46%
Neutral	2.095	21,24%
Positive	7.357	74,60%
Total	9.862	100%

Tabla 5. Análisis de sentimiento con Azure / Fuente: Twitter / Azure / Elaboración propia.

Después de exponer los datos del análisis de sentimiento de las cuatro bases de datos correspondientes a los diferentes *hashtags* podemos ver que en todas las etiquetas los *tweets* con una valoración positiva son mayoría respecto a los valorados como neutros y negativos.

Si nos paramos a comparar los porcentajes de cada valoración inter relacionándolos con los diferentes *hashtags* vemos que:

-La valoración negativa en las cuatro bases de datos se encuentra entre el 4,27% y el 5,36%.

-La valoración neutra en las cuatro bases de datos se encuentra entre el 18,17% y el 21,70%.

-La valoración positiva entre las cuatro bases de datos se encuentra entre el 73,29% y 76,36%.

Con esto vemos que entre las diferentes etiquetas se guarda una relación de coincidencia en los tres niveles de valoración.

7.2. Resultado del análisis de sentimiento manual

Como se ha explicado en el apartado de los objetivos uno de ellos es la detección de errores en los análisis para la posterior mejora de los *softwares*. Además, en la metodología explicamos que se haría una análisis de sentimiento manual con el *hashtag* *#LasFinalesDeTodaLaVida* ya que es la base de datos más pequeña (732 *tweets*) pero sigue siendo una muestra suficiente para detectar errores en los análisis de sentimiento.

Después de realizar el análisis de sentimiento manual los resultados que hemos obtenido han sido los siguientes:

Análisis de sentimiento manual #LasFinalesDeTodaLaVida

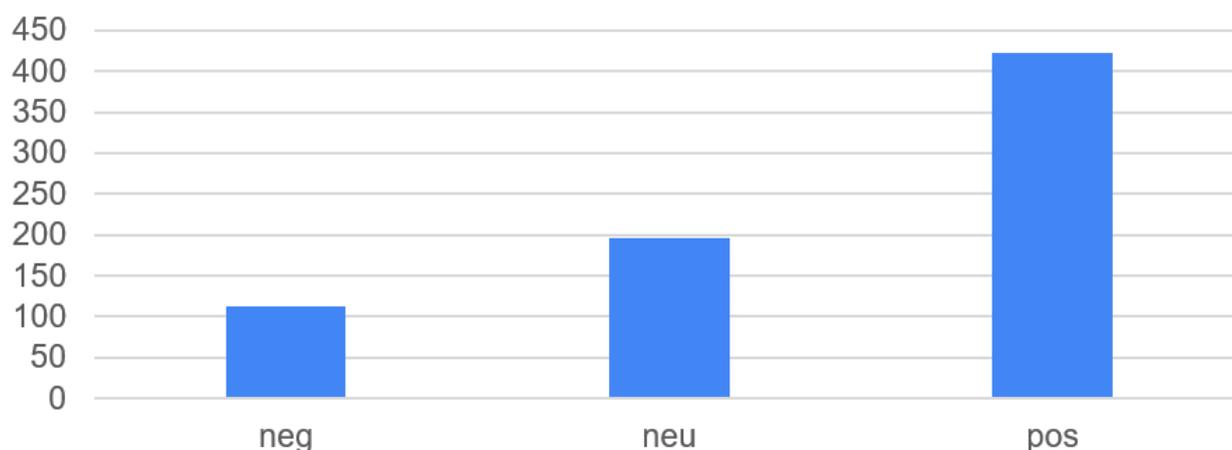


Gráfico 6. Resultados análisis de sentimiento manual de #LasFinalesDeTodaLaVida / Fuente: Twitter / Elaboración propia.

Concretamente, los datos expuestos en el gráfico de barras son los siguientes:

Sentimiento	Total	Porcentaje
Negative	112	15,30%
Neutral	196	26,78%
Positive	423	57,79%
Total	732	100%

Tabla 6. Resultados análisis de sentimiento manual de #LasFinalesDeTodaLaVida / Fuente: Twitter / Elaboración propia.

El objetivo de este análisis de sentimiento manual es comparar los resultados con los obtenidos de forma automática con el complemento de *Microsoft Excel*, *Azure*. Este análisis manual se ha realizado únicamente de la base de datos perteneciente al *hashtag* #LasFinalesDeTodaLaVida, ya que es la base de datos más pequeña (732 *tweets*) y es considerada de un tamaño correcto para que la comparación sea válida.

Los resultados obtenidos con *Azure* fueron los siguientes:

Sentimiento	Total	Porcentaje
Negative	40	5,46%
Neutral	133	18,17%
Positive	559	76,36%
Total	732	100%

Tabla 7. Análisis de sentimiento con Azure de #Las Finales De Toda La Vida / Fuente: Twitter / Azure / Elaboración propia.

Y los obtenidos de forma manual han sido estos:

Sentimiento	Total	Porcentaje
Negative	112	15,30%
Neutral	196	26,78%
Positive	423	57,79%
Total	732	100%

Tabla 8. Resultados análisis de sentimiento manual de #LasFinalesDeTodaLaVida / Fuente: Twitter / Elaboración propia.

Sentimiento	Diferencia
Negative	+9,84%
Neutral	+8,67%
Positive	-18,57%

Tabla 9. Comparación de análisis manual con Azure en #LasFinalesDeTodaLaVida / Fuente: Twitter / Elaboración propia.

Esta diferencia en los resultados, entre el análisis manual y el automático, puede deberse a una falta de contexto por parte del software, y que nosotros como persona y aficionado al evento analizado sí que podemos llegar a comprender e incluso llegar a poder valorar de forma positiva o negativa *tweets* que realmente tienen muy poco contexto, pero que por nuestro bagaje llegamos a comprender y valorar sin problema.

Por otro lado, comprender el lenguaje natural por un *software* sin errores es difícil. Esto se debe a las características del propio lenguaje, como por ejemplo la ironía o el sarcasmo, que dificultan la correcta identificación del mensaje.

Este problema se podría solucionar con la inteligencia artificial (IA) de una forma muy sencilla de entender. Todos los *tweets* extraídos con la herramienta *Twitter Archiver* llevan consigo una fecha y un hora de publicación. Usando estos datos la IA podría recopilar todos esos *post* de forma cronológica y podría crear un relato que sirva de contexto para valorar *tweets* que de forma individualmente podrían llegar a ser confusos para el *software* de *Azure*.

A continuación podemos ver algunos ejemplos de los *tweets* que han cambiado de valoración al pasar por análisis manual.

Tweet	Azure	Manual
¡Enorme remontada y victoria de Boston en el Game 1 de las Finales! Los Celtics completan un último cuarto excepcional con un 9 de 12 en triples y un parcial total de 40-16 para robar el factor cancha a los Warriors. #NBAFinals #LasFinalesDeTodaLaVida	Neutral	Positive
“La noche de los picorcitos” @ADaimiel 21h : 1º partido Final @ACBCOM @FCBbasket vs @RMBaloncesto 3h : 5º partido Final @NBA @celtics vs @warriors #DormirEsDeCobardes #LasFinalesDeTodaLaVida	Positive	Neutral
#LasFinalesDeTodaLaVida A Green le van a permitir todo? Lo llega a hacer Smart y le cae técnica	Positive	Negative
#LasFinalesDeTodaLaVida aunq Warriors no ganase las Finales yo le daba el MVP a Curry porq es tremendo lo q esta haciendo	Neutral	Positive
#LasFinalesDeTodaLaVida como me gustan las finales años y años viéndolas con vosotros... infinitamente gracias @ADaimiel formas parte de mi vida desde la adolescencia !!!	Neutral	Positive
#LasFinalesDeTodaLaVida con 14 de 23 en tiros libres y 18 pérdidas el partido lo pierdes tú solito @BostonCeltics	Neutral	Negative

#LasFinalesDeTodaLaVida El arbitraje favorece una vez más a Boston ! #NBAFinals2022	Positive	Negative
#LasFinalesDeTodaLaVida es imposible ganar con este arbitraje. A la cama, no merece la pena seguir aguantando más injusticia por hoy. @ADaimiel se va a la cama contento esta noche.	Positive	Negative

Tabla 10. Ejemplo de valoración de tweets entre análisis manual y automático / Fuente: Twitter / Azure / Elaboración propia.

Estos cambios de valoración, dependiendo del tipo de análisis de sentimiento, vienen sin duda por diferentes motivos, estos pueden ser faltas de ortografía y malas comprensiones de léxicos, pero sin duda alguna el mayor motivo es la falta de contexto que sufre el software.

7.3. Resultados del análisis de clusters

Como hemos explicado anteriormente los datos que vamos a exponer aquí son las tablas tituladas 'Results by clusters' y 'Clusters centroids', aunque la aplicación *XLStat* nos permite ver muchos más datos (los cuales estarán disponibles en bruto en el Anexo de este documento). La exposición de los datos la haremos a continuación por *hashtags*.

En la etiqueta *#NBAPlayoffs* la aplicación *XLStat* nos ha arrojado los siguientes datos:

Clase	1	2	3	4	5
Number of objects by cluster	2894	191	194	50	44
Suma de los pesos	2894	191	194	50	44
Within-cluster variance	#####	#####	#####	#####	#####
Distancia mínima al centroide	81,261	2253,753	27266,615	95519,637	3502,849
Distancia media al centroide	5074,190	39774,938	86643,170	336071,143	38809,867
Distancia máxima al centroide	45886,641	110207,469	1257140,601	2665904,722	166177,925

Tabla 11. Resultados por clusters de XLStat de #NBAPlayoffs / Fuente: Twitter / XLSTAT / Elaboración propia.

Clase	Retweets	Favorites	Followers	Follows	Suma de los pesos	Within-cluster variance
1	0,444	1,912	3780,751	909,947	2894,000	75367396,035
2	1,518	9,932	101446,984	2825,403	191,000	2109230520,898
3	19,046	221,722	711981,902	1261,722	194,000	37969789745,750
4	1,240	8,020	3489914,280	3887,960	50,000	334695669338,859
5	11,682	71,750	370438,795	931,545	44,000	3904375084,555

Tabla 12. Resultado de los clusters centroides de XLSTAT de #NBAPlayoffs / Fuente: Twitter / XLSTAT / Elaboración propia.

Para analizar los resultados correspondientes a la etiqueta #NBAPlayoffs vamos a hacer referencia a los datos de la tabla 11 (Tabla 11) y la tabla 12 (Tabla 12).

La primera de estas tablas nos indica que los usuarios que *postean* con esta etiqueta se pueden dividir en cinco tipos de usuarios (*clusters*). Y vemos que el *cluster* más grande es el primero y con mucha diferencia respecto al resto, formado por un total de 2.894 usuarios.

Por la experiencia que tenemos con este *software* sabemos que pueden existir *clusters* formados por un solo usuario, así que, buscando en la información que no proporciona *XLStat* (la cual solo se podrá ver en el anexo), hemos visto que no es el caso. Todos los *clusters* están formados por diferentes usuarios de *Twitter*.

La segunda de las tablas nos refleja los datos de actividad media entre los diferentes usuarios, ya que a la hora de realizar el análisis (como explicamos en la metodología) escogimos las columnas de ‘seguidos’, ‘seguidores’, ‘retweets’ y ‘favoritos’.

Del primer *cluster* podemos ver que está formado por usuarios con poco *feedback* en sus *tweets*. Estos obtienen una media de 0,4 *retweets* y casi 2 (1,9) favoritos. Vemos que son cuentas que tienen bastantes seguidores, una media de casi 3.800 seguidores y que siguen a bastantes menos cuentas, cerca de 1.000.

Del segundo *cluster* podemos ver que gana un poco más de interactividad respecto al primero, ya que obtiene casi un *retweet* más (1,51) y casi 8 favoritos más (9,93). Vemos también que el tamaño de estas cuentas es mucho mayor respecto a los seguidores y seguidos. Son usuarios con más de 100.000 seguidores de media (101.446,984) y que siguen a muchísimas menos cuentas, alrededor de los 2.800 usuarios seguidos (2.825,403).

Del tercer *cluster* vemos que son usuarios los cuales ganan mucha más repercusión con sus *tweets* ya que obtienen una media de 19 *retweets* y más de 220 favoritos (221,72). Son usuarios que consiguen ganar en seguidores al anterior *cluster* con cerca de 772.000 (711.981,902) y que también sigue a menos usuarios con una media de 1.261,722.

En el cuarto *cluster* vemos un fenómeno un tanto raro. Este corresponde a usuarios con poca interacción, obtienen poco más de un *retweet* (1,24) y unos 8 favoritos. Estas cifras nos extrañan ya que es el *cluster* que más seguidores tiene con una media de 3.489.914,280 y que en comparación únicamente sigue a una media cercana a las 4.000 cuentas (3.887,96).

En el último *cluster*, el quinto, podemos ver que son usuarios con una considerable cantidad de *feedback*, ya que consigue una media de 11,68 *retweets* y unos 72 favoritos por *tweet* (71,75). Su cifra de seguidores desciende considerablemente en relación al anterior *cluster*, con una media 370.438,795 y su cifra de cuentas seguida sigue la misma tendencia con 931,545 de media.

En la *hashtag* #TiempoDePlayoffs la aplicación *XLStat* nos ha arrojado los siguientes datos:

Clase	1	2	3	4	5
Number of objects by cluster	156	763	17	151	20
Suma de los pesos	156	763	17	151	20
Within-cluster variance	12761966,204	337628,059	310220025,544	177737369,685	132484534,932
Distancia mínima al centroide	337,578	36,037	3234,218	2990,863	2609,262
Distancia media al centroide	2612,459	428,060	9455,196	6608,309	10242,081
Distancia máxima al centroide	15459,753	3137,881	62280,008	51307,690	21143,763

Tabla 13. Resultados por clusters de XLStat de #TiempoDePlayoffs / Fuente: Twitter / XLSTAT / Elaboración propia.

Clase	Retweets	Favorites	Followers	Follows	Suma de los pesos	Within-cluster variance
1	0,494	3,724	2908,756	2199,244	156,000	12761966,204
2	0,060	0,451	262,010	416,448	763,000	337628,059
3	8,529	63,706	126214,647	391,765	17,000	310220025,544
4	8,636	52,066	312353,510	240,417	151,000	177737369,685
5	1,250	14,700	44930,550	3443,600	20,000	132484534,932

Tabla 14. Resultado de los clusters centroides de XLSTAT de #TiempoDePlayoffs / Fuente: Twitter / XLSTAT / Elaboración propia.

Para analizar los resultados correspondientes al *hashtag* #TiempoDePlayoffs vamos a hacer referencia a los datos de la tabla 13 (Tabla 13) y la tabla 14 (Tabla 14).

En esta primera tabla (Tabla 13) los datos nos muestran que existen también cinco *clusters* en esta etiqueta. Siendo el segundo el mayor de estos, formado por un total de 763 usuarios. En este análisis tampoco encontramos *clusters* formados por un único usuario.

Como ya hemos explicado anteriormente, la segunda tabla (Tabla 14) nos muestra el grado de interactividad medio de los usuarios que forman los *clusters*.

El primer *cluster* obtiene muy poca actividad en cuanto al *feedback* de sus *tweets*, ya que obtiene menos de un *retweet* de media (0,49), aunque si consigue casi cuatro favoritos por post (3,72). Los usuarios que conforman este primer cluster siguen a casi 3.000 cuentas (2.908,76) y son seguidos por cerca de 2.200 usuarios (2.199,24).

El segundo *cluster* tiene los peores datos en todos los apartados respecto al resto de *clusters*. Este no suele obtener ningún *retweet* (0,06) y obtiene una media de 0,45 favoritos. Es el *cluster* con menos seguidores de media (262) y sigue de media a unos 416,45 usuarios.

El tercer *cluster* mejora todos los datos, ya que obtiene de media casi 9 *retweets* (8,53) y recibe de media 63,71 favoritos por *post*. Además, ya tiene una gran número medio de seguidores, 126.214,68; y, en comparación, únicamente sigue a cerca de 400 cuentas (391,77).

El cuarto *cluster* muestra unas estadísticas de *feedback* muy parecidas al anterior casi 9 *retweets* (8,64) y 52 favoritos de media (52,06) por *tweet*. En cuanto a seguidores es el *cluster* que más tiene de media, 312.353,51 y también es el que menos usuarios sigue, ronda los 240 seguidos (240,42).

El último *cluster*, el quinto, rebaja sus tasas de interacción respecto al tercer y el cuarto *cluster*. Este obtiene alrededor de un *retweet* (1,25) y casi 15 favoritos de media (14,7) por *post*. En cuanto a usuarios seguidos y seguidores, es el tercer

cluster con más seguidores, cerca de 45.000 (44.930,55), y a su vez es el que más usuarios sigue con 3.443,6 de media.

En la etiqueta *#NBAFinals* la aplicación *XLStat* nos ha arrojado los siguientes datos:

Clase	1	2	3	4	5
Number of objects by cluster	3939	218	228	229	70
Suma de los pesos	3939	218	228	229	70
Within-cluster variance	#####	2532753174,327	1289903913,560	#####	#####
Distancia mínima al centroide	199,951	3086,012	2046,665	7508,668	9523,492
Distancia media al centroide	8410,858	29766,980	27242,951	279987,934	430318,059
Distancia máxima al centroide	68070,024	166534,449	74124,586	1214657,773	2501171,872

Tabla 15. Resultados por clusters de XLStat de #NBAFinals / Fuente: Twitter / XLSTAT / Elaboración propia.

Clase	Retweets	Favorites	Followers	Follows	Suma de los pesos	Within-cluster variance
1	0,493	2,613	6433,406	981,664	3939,000	171651592,241
2	9,702	74,596	370077,968	3320,358	218,000	2532753174,327
3	2,732	19,364	138672,518	4545,820	228,000	1289903913,560
4	17,406	209,293	851493,266	1624,865	229,000	157501780041,334
5	1,786	15,243	3660513,043	3716,786	70,000	653531090386,918

Tabla 16. Resultado de los clusters centroides de XLSTAT de #NBAFinals / Fuente: Twitter / XLSTAT / Elaboración propia.

Para analizar los resultados correspondientes al *hashtag* *#NBAFinals* vamos a hacer referencia a los datos de la tabla 15 (Tabla 15) y la tabla 16 (Tabla 16).

Esta primera tabla (tabla 15) nos muestra que, como hemos visto ahora, esta etiqueta está conformada por cinco *cluster*, y ninguno de estos está formado por un solo usuario. De estos cinco *cluster* el primero es el más grande al constar de un total de 3939 usuarios.

En cuanto a la segunda tabla (tabla 16) podemos ver que el primer *cluster* es el que menos interactividad tiene con menos de *retweet* (0,5) y casi 3 favoritos de media (2,61) por *tweet*. Vemos también que es el *cluster* con menos seguidores de media (6.443,41) y con menos de mil usuarios seguidos (981,66).

El segundo *cluster* es uno de los que más interactividad tiene, concretamente el segundo. Este obtiene de casi 10 *retweets* (9,7) y cerca de 75 favoritos (74,60) de media por *post*. Los usuarios que forman este *cluster* tienen 370.077,97 seguidores y siguen a 3.320,358 usuarios de media.

El tercer *cluster* baja sus tasas de interactividad en comparación con el anterior. Este casi llega a los 3 *retweets* (2,73) y a los 20 favoritos (19,36) de media por *tweet*. También disminuye su media de seguidores con 138.672,518 y la de usuarios seguidos con 4.545,82.

El siguiente *cluster*, el cuarto, es el que más interactividad tiene entre todos ellos, pues consigue unos 17 *retweets* (17,41) y unos 210 favoritos (209,293) de media por *post*. En cuanto a seguidores, es el que más seguidores tiene de media, con 851.493,27 y el que menos usuarios sigue con 1.624,87.

El quinto *cluster* no consigue tanta tasa de interacción, de hecho, es el penúltimo en estos apartados. Consigue casi 2 *retweets* (1,79) y cerca de los 15 favoritos (15,24) de media por cada *tweet*. Una cosa a destacar es su número medio de seguidores, ya que alcanza los 3.660.513,04 y únicamente sigue a una media de 3.716,79.

En la *hashtag* #LasFinalesDeTodaLaVida la aplicación XLStat nos ha arrojado los siguientes datos:

Clase	1	2	3	4	5
Number of objects by cluster	67	636	15	1	13
Suma de los pesos	67	636	15	1	13
Within-cluster variance	82222,226	4871115,988	31346666,467	0,000	#####
Distancia mínima al centroide	39,557	70,261	1367,798	0,000	2714,659
Distancia media al centroide	228,656	1220,046	2700,275	0,000	7158,234
Distancia máxima al centroide	1225,577	17371,160	20235,655	0,000	12757,753

Tabla 17. Resultados por clusters de XLStat de #LasFinalesDeTodaLaVida / Fuente: Twitter / XLSTAT / Elaboración propia.

Clase	Retweets	Favorites	Followers	Follows	Suma de los pesos	Within-cluster variance
1	25,448	146,731	309652,836	200,000	67,000	82222,226
2	0,079	0,807	825,918	813,299	636,000	4871115,988
3	15,067	97,933	124439,533	401,733	15,000	31346666,467
4	27,000	480,000	615445,000	201,000	1,000	0,000
5	4,615	34,308	42830,154	1909,692	13,000	67585616,359

Tabla 18. Resultado de los clusters centroides de XLSTAT de #LasFinalesDeTodaLaVida / Fuente: Twitter / XLSTAT / Elaboración propia.

Para analizar los resultados correspondientes al *hashtag* #LasFinalesDeTodaLaVida vamos a hacer referencia a los datos de la tabla 17 (Tabla 17) y la tabla 18 (Tabla 18).

Si nos fijamos en la primera tabla (tabla 17), al igual que el resto este *hashtag* está formado por cinco *clusters*. Además, de los cinco *clusters* el más grande es el segundo con un total de 636 usuarios

Aún así, esta tabla nos muestra una diferencia, y es que, esta vez sí que hay uno de ellos compuesto por un solo usuario, concretamente el cuarto cluster lo compone el usuario *@vamos* que es la cuenta oficial del canal de televisión *Vamos* de *Movistar+*, canal por el cual se retransmiten los partidos pertenecientes a los *Playoffs* de la NBA..

En cuanto a la segunda tabla (tabla 18) podemos ver que el primer *cluster* nos muestra una gran tasa de interacción con unos 25 *retweets* (25,45) y casi 150 favoritos (146,73) de media por *tweet*. Su alto nivel de *feedback* puede deberse a su gran número de seguidores, con 309.652,83 de media. Las cuentas que forman este primer *cluster* solo siguen a 200 usuarios de media.

El segundo *cluster* es el que menos tasa de interacción muestra, con menos de un *retweet* y un favorito de media, respectivamente 0,08 y 0,81. También es el *cluster* con menos seguidores, 825,92 de media y, a su vez, es el segundo que más usuarios sigue, con alrededor de 800 (813,30).

El siguiente aumenta sus tasas de interacción, posicionándose como el tercer *cluster* con más interacción, ya que consigue unos 15 *retweets* de media (15,07) y casi 100 favoritos (97,93). También es el tercero con más seguidores, puesto que tiene una media de 124.439,53 y los usuarios que lo conforman siguen a cerca de 400 cuentas (401,73).

El cuarto *cluster*, recordamos que está formado por una única cuenta y un único *tweet*, consiguió 27 *retweets* y 480 favoritos. La cuenta de *@vamos* tenía en el momento del análisis 615.445 seguidores y seguía a un total de 201 usuarios.

Por último, el quinto *cluster* obtuvo 4,62 *retweets* y 3434,31 favoritos de media. En cuanto a seguidores y seguidos, los usuarios que conforman este *cluster* tenían casi 43.000 seguidores (42.830,15) y rondaban los 2.000 usuarios seguidos (1.909,70).

8. Conclusiones

Tras la exposición de los resultados y su correspondiente análisis hemos extraído las siguientes conclusiones:

8.1. Conclusiones del análisis de sentimiento

- La NBA es un producto de nicho, por lo que solo *tuitean* sobre este tema usuarios a los que les gusta el producto, de ahí que al realizar el análisis de sentimiento automático con *Azure* la tasa de valoraciones positivas obtenidas sea de entre un 73,29% y un 76,36%.
- Tras realizar el análisis de sentimiento manual y compararlo con el automático todas las valoraciones (negative, neutral y positive) han sufrido grandes variaciones. Esto lo achacamos a la falta de contexto que sufre el software pero que en cambio nosotros sí tenemos.

8.2. Conclusiones del análisis de *clusters*

- Referente a la etiqueta *#NBAPlayoffs* pensamos que el tercero y el quinto pueden estar formados por usuarios relevantes en el sector, es decir, *influencers*. Pensamos esto por sus altos números de seguidores y sus tasas de interacción. Obviamente hay que tener en cuenta que serían *influencers* de diferente tamaño, el tercero sería el que estaría compuesto por *influencers* de mayor tamaño. También creemos que el cuarto *cluster* puede estar formado por usuarios *bots* o por usuarios que compran seguidores, ya que llama la atención la poca tasa de interacción que obtiene con casi 3.500.00 seguidores de media. El resto de *clusters* consideramos que sin usuarios normales.
- En cuanto al *hashtag* *#TiempoDePlayoffs* encontramos que los *clusters* que están formados por grandes *influencers* son el tercero y el cuarto, ya que son usuarios con cerca de 125.000 y 310.000 usuarios respectivamente. Además, consideramos que el quinto está formado por *microinfluencers*, ya que tiene muchos menos seguidores que los anteriores dos *clusters*, pero sigue siendo una cantidad considerable. Como consecuencia, su tasa de interacción baja, pero sigue siendo grande.

- Sobre la etiqueta *#NBAFinals* concluimos que hay un *cluster*, concretamente el cuarto, que podemos considerar que está formado por *influencers*. A su vez, podemos creer que el segundo y el tercer *cluster* están conformados por *microinfluencers*. Y por último, encontramos, al igual que en etiqueta *#NBAPlayoffs*, un *cluster* que parece estar formado por usuarios que compran seguidores, ya que tiene el mayor número medio de seguidores (3.660.513,04), pero su tasa de seguidores es la más baja comparada con el segundo, el tercer y el cuarto *cluster*.
- Por último, en la etiqueta *#LasFinalesDeTodaLaVida* encontramos tres *clusters* que podemos considerar que están formados por *influencers*, concretamente nos referimos al primero, tercero y cuarto *cluster*. Los otros dos *clusters* restantes, el segundo y el quinto, parecen ser simples usuarios, puesto que sus niveles de interacción van acorde a su volumen de seguidores. Siendo más concretos, al quinto *cluster* podemos creer que está formado por *microinfluencers*.

9. Relación del trabajo realizado en relación a los estudios cursados

Durante el curso académico 21/22 cursé los estudios de Máster en Comunicación Transmedia. Dentro de su programa curricular se encontraba la asignatura de Audiencia participativa y social. Durante el transcurso de la misma, estudiamos lo que se ha realizado en este trabajo de investigación, el análisis de sentimientos automático y el análisis de *clusters*, pero en trabajos más pequeños y sin repercusión académica para estas metodologías.

Con este Trabajo Final de Máster se ha buscado implementar estas técnicas de análisis de usuarios y comunidades en redes sociales para llevarlas a cabo en estudios más grandes como es este, y así poder encontrar fallos en ambas metodologías para que sus campos de estudio sigan avanzando en la línea correcta. Además, esto también me ayuda a mi particularmente en conocer metodologías más profesionales las cuales podré aplicar en mi futuro laboral teniendo así lo que considero un punto de diferenciación respecto a algunos profesionales del sector de la comunicación, las redes sociales, el marketing, etc.

10. Referencias

11.1. Bibliografía

- Arcila-Calderón, C., Ortega-Mohedano, F., Jiménez-Amores, J., y Trullenque. (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), 973-982.
- Baviera, T. (2017).** Técnicas para el análisis del sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength. *Dígitos*, 1(3), 33-50.
- C. Honey y S.C. Herring (2009).** Beyond Microblogging: Conversation and Collaboration via Twitter
- Fundeú BBVA, & Tascón, M. (2012).** *Escribir en internet: Guía para los nuevos medios y las redes sociales (Rústica)* (2.^a ed.). Galaxia Gutenberg.
- Gallardo Campos, M. (2009).** Aplicación de técnicas de clustering para la mejora del aprendizaje (Master's thesis).
- García-Teodoro, P., J. Díaz-Verdejo, G. Maciá-Fernández and E. Vázquez (2009).** Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, 2009, 28(1), 18-28.
- Martínez-Fresneda Osorio, H., y Sánchez Rodríguez, G. (2022).** La influencia de Twitter en la agenda setting de los medios de comunicación. *Revista de Ciencias de la Comunicación e Información*, 27, 1-21.
- Medhat, W., Hassan, A., & Korashy, H. (2014).** Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.

Porter, M. E. (1999). Ser competitivo: Fronteras en expansión. *Harvard Deusto business review*, (91), 34-36.

Scott, S., & Matwin, S. (1998). Using Lexical Knowledge in Text Classification. Available as University of Ottawa Department of Computer Science Technical Report TR-98-03.

Vásquez, A. C., Quispe, J. P., & Huayna, A. M. (2009). Procesamiento de lenguaje natural. *Revista de investigación de Sistemas e Informática*, 6(2), 45-54.

11.2. Webliografía

A. (2021, 2 noviembre). *Language Detection language support - Azure Cognitive Services.* Microsoft Docs. <https://docs.microsoft.com/en-us/azure/cognitive-services/language-service/language-detection/language-support>

Atribus. (2022, 30 agosto). *¿Qué es el SOCIAL LISTENING? Ventajas en 2022.* <https://atribus.com/social-listening/>

Cloud Natural Language|. Google Cloud. <https://cloud.google.com/natural-language?hl=es>

Frequently asked questions. (2005). WordNet. <https://wordnet.princeton.edu/frequently-asked-questions>

NBA en Movistar Plus+ (@MovistarNBA) /. (2022). [Tweet]. Twitter. <https://twitter.com/MovistarNBA>

NBA (@NBA) /. (2022). [Tweet]. Twitter. <https://twitter.com/NBA>

Parra, A. (2021, 15 enero). *Análisis de sentimiento. Qué es y cómo realizarlo.*

QuestionPro.

[https://www.questionpro.com/blog/es/herramienta-de-analisis-de-sentimientos
/#Usos_de_la_herramienta_de_analisis_de_sentimiento](https://www.questionpro.com/blog/es/herramienta-de-analisis-de-sentimientos/#Usos_de_la_herramienta_de_analisis_de_sentimiento)

Sentiment Analysis in Excel | MeaningCloud.

<https://www.meaningcloud.com/developer/excel-addin/doc/3.5/sentiment-analysis>

SentiStrength - sentiment strength detection in short texts - sentiment analysis, opinion mining. <http://sentistrength.wlv.ac.uk/#About>

Servicio de análisis de sentimiento | Shaip.

<https://es.shaip.com/solutions/sentiment-analysis-services/#>

Speranza, M. F. *WordNet Domains.* <https://wndomains.fbk.eu/wnaffect.html>

11. Anexos

En este apartado se puede acceder a las cuatro bases de datos usadas para los análisis y a los informes aportados por *Atribus*.

[ENLACE](#)