



Universidad
Zaragoza

Trabajo Fin de Máster

Análisis de las comunicaciones en Twitter de la
Universidad de Zaragoza: un modelo empírico

Analysis of the Twitter communications of the
University of Zaragoza: an empirical model

Autor

José Miguel Pina Pérez

Director

Sergio Ilarri Artigas

Facultad de Filosofía y Letras

Máster en Consultoría de Información y Comunicación Digital 2020/2021

AGRADECIMIENTOS

En el desarrollo de este trabajo han sido fundamentales las orientaciones realizadas por mi tutor Sergio Ilarri. Además de reconocer su esfuerzo, y valorar su confianza, también quería transmitirle mi agradecimiento por acercarme al mundo de la minería de datos y contagiarme por la ilusión de aprender unas herramientas que antes de empezar este trabajo desconocía.

También me parece justo reconocer la labor realizada por aquellas personas anónimas y organizaciones que ofrecen de manera libre y gratuita contenidos formativos, códigos de programación, programas informáticos y otras herramientas. Esta labor es fundamental para que muchos investigadores, empresas e instituciones puedan avanzar y contribuir entre todos a la construcción de una verdadera sociedad del conocimiento.

No quiero finalizar sin recordar el apoyo de mi mujer y de mi hija, siempre fundamentales para encontrar la motivación y energía necesarias para sacar adelante cualquier proyecto vital.

RESUMEN

La utilización de redes sociales como Twitter es fundamental en las comunicaciones institucionales. Analizando una muestra de 53.783 tuits publicados por 55 cuentas vinculadas a la Universidad de Zaragoza, en este estudio se propone y se evalúa un modelo empírico que explica cuáles son los factores que más influyen en la efectividad de los tuits. Los factores analizados abarcan desde el prestigio asociado a la cuenta hasta los atributos tangibles y emocionales presentes en el mensaje. Mediante diferentes herramientas como Python, KNIME, VBA e IBM SPSS, se demuestra que el número de *likes* y retuits de un tuit determinado depende del número de seguidores de una cuenta, de la temática de la cuenta, de la temática del tuit, de la existencia de elementos audiovisuales, *hashtags* y menciones, de la legibilidad de la información y del sentimiento que evoca el tuit a través del texto y de sus emojis.

Palabras clave: Twitter, Universidad de Zaragoza, comunicación, análisis del sentimiento

ABSTRACT

Social networks such as Twitter are essential for institutional communication. Analysing a sample of 53,783 tweets published by 55 accounts from the University of Zaragoza, this study proposes and tests an empirical model that explains the factors that most influence the effectiveness of tweets. These factors include the account's perceived prestige and the tangible and emotional attributes included in the message. Using tools such as Python, KNIME, VBA and IBM SPSS, it was found that the number of likes and retweets received by any tweet depends on the account's number of followers, the topic of the account, the topic of the tweet, whether it features audio-visual elements, hashtags and mentions, the readability of the information and the feelings that the tweet text and emojis evoked.

Keywords: Twitter, Universidad de Zaragoza, communication, sentiment analysis

ÍNDICE GENERAL

1. Introducción.....	7
2. Marco teórico.....	9
2.1 Twitter como red social.....	9
2.2 Twitter en la investigación académica	11
2.3 Efectividad de las comunicaciones en Twitter. Modelo propuesto.....	13
3. Metodología	17
3.1 Herramientas utilizadas	18
3.2 Contexto de estudio.....	18
3.3 Selección de la muestra.....	23
3.4 Medición de las variables.....	27
4. Resultados	28
4.1. Identificación de los elementos del tuit.....	29
4.2. Análisis semántico I: temática de los tuis	31
4.3. Análisis semántico II: sentimiento del texto	37
4.4. Análisis semántico III: sentimiento de los emojis.....	40
4.5. Análisis de la legibilidad.....	40
4.6. Contraste de hipótesis.....	41
5. Conclusiones, implicaciones y trabajos futuros.....	44
Referencias.....	47
Anexo I. Diccionario de términos técnicos	53
Anexo II. Compromisos de calidad e indicadores del Gabinete de Comunicación	52
Anexo III. Organismos de la UZ con cuentas en redes sociales	55
Anexo IV. Código para extraer tuits de varias cuentas simultáneamente	58
Anexo V. Código para construir y evaluar el modelo de clasificación	59

TABLAS Y FIGURAS

Tabla 2.1. Uso de las redes sociales en las organizaciones.....	13
Tabla 2.2. Fórmulas de legibilidad clásicas	16
Tabla 3.1. Presencia en redes sociales de organismos vinculados a la UZ.....	21
Tabla 3.2. Métricas de las cuentas de la UZ en redes sociales.....	22
Tabla 3.3. Población de cuentas utilizadas en la investigación.....	24
Tabla 3.4. Muestra de cuentas utilizadas en la investigación.....	26
Tabla 3.5. Medición de variables	28
Tabla 4.1. Resultados de los algoritmos de clasificación.....	36
Tabla 4.2. Estimación de los modelos de regresión jerárquica	42
Figura 2.1. Imagen de Twitter.....	10
Figura 2.2. Redes sociales más utilizadas	11
Figura 2.3. Evolución del número de publicaciones sobre redes sociales	12
Figura 2.4. Modelo sobre la efectividad de las comunicaciones en Twitter.....	14
Figura 3.1. Mapa de redes sociales de la UZ	20
Figura 3.2. Vista de Python durante la descarga de información	23
Figura 4.1. Código para identificar los emojis de los tuits	29
Figura 4.2. Código para extraer las URL extendidas de los tuits.....	30
Figura 4.3. Proceso utilizado en el análisis semántico.....	31
Figura 4.4. Código para extraer los <i>hashtags</i> del texto.....	32
Figura 4.5. Código para limpiar los caracteres no alfabéticos	33
Figura 4.6. Flujo definido en KNIME para procesar el texto	34
Figura 4.7. Código para construir y evaluar el modelo predictivo (extracto).....	35
Figura 4.8. Código para predecir la categoría de los tuits.....	36
Figura 4.9. Código de análisis de sentimiento con VADER.....	38
Figura 4.10. Código de análisis de sentimiento con TextBlob	39
Figura 4.11. Comparativa entre los resultados de TextBlob y VADER.....	40
Figura 4.12. Función para contar sílabas	41

1. Introducción

El surgimiento de las redes sociales en la década de los 2000 supuso una auténtica revolución en los patrones de comunicación de las personas y sus relaciones con las organizaciones (Qualman, 2012; Wilcox et al., 2012). En la actualidad, plataformas como Twitter, Facebook o Instagram se han convertido en el altavoz principal para muchas marcas y personas a nivel individual, las cuales pueden trabajar su reputación con un coste por impacto residual.

Lejos de ser un fenómeno pasajero, el mundo de las redes sociales sigue evolucionando e incorporando nuevos actores, como Twitch y TikTok, que acaparan la atención de los usuarios más jóvenes. Asimismo, empresas de sectores concretos como el turismo también se esfuerzan en ofrecer en sus plataformas (ej. Booking, TripAdvisor) elementos de interacción propios de las redes sociales, con el fin de que sus usuarios puedan dialogar, compartir experiencias y crear comunidades de marca (Zhang et al., 2017).

Desde su nacimiento, las redes sociales se han considerado una herramienta imprescindible para la comunicación y las relaciones públicas (Wilcox et al., 2012). Estas redes se han convertido en los canales de información, comunicación y socialización favoritos de millones de individuos, siendo la única manera para muchas marcas de contactar con su público. De acuerdo con el estudio anual del *Interactive Advertising Bureau* (IAB, 2020), el 72% de los usuarios sigue cuentas de marcas en las redes sociales, de los que la mitad aproximadamente indican que utilizan mucho o bastante las redes con este fin.

En definitiva, la importancia cuantitativa y cualitativa del fenómeno de las redes sociales obliga a un proceso de reflexión en las organizaciones sobre cuál debe ser su presencia en dichas plataformas, y qué tipo de estrategias pueden resultar más adecuadas para conseguir una mayor visibilidad y reputación online. Este Trabajo Fin de Máster (TFM) se centra en particular en la red social Twitter, un canal que se considera muy útil para dialogar con el público objetivo de una organización y entablar relaciones duraderas (Wang y Yang, 2020). En la literatura académica, numerosos trabajos han puesto la lupa sobre esta red social, tratando de estudiar cómo se desarrollan las publicaciones en este canal y cómo los consumidores responden a dichas publicaciones (Zimmer y Proferers, 2014; McCormick et al., 2017; Stewart y Quan-Haase, 2017; Karami et al., 2020; Noor et al., 2020). Sin embargo, la mayoría de los estudios existentes se centran en la vertiente del consumidor, habiendo una notable carencia de estudios empíricos que planteen modelizaciones o aproximaciones globales que permitan conocer a las empresas qué factores influyen más en la efectividad de sus mensajes en esta plataforma.

En concreto, el *objetivo principal* de este trabajo es proponer y evaluar un modelo empírico que, partiendo de la literatura previa, identifique claramente cuáles son los factores que determinan la efectividad de un *tuit*. Tal y como iremos desarrollando en los siguientes epígrafes esta efectividad será medida en términos del número de *likes* y *retuits* que recibe dicho tuit, mientras que como factores explicativos analizaremos la reputación de la cuenta, los propios elementos incluidos en dicho tuit (*hashtags*, menciones, etc.), la temática y legibilidad de la información, así como la emoción generada. Para poder medir estos aspectos, se pretende utilizar diferentes herramientas y lenguajes de programación, siendo de especial interés mostrar la aplicación de las técnicas basadas en el análisis del sentimiento para determinar la valencia emocional de los mensajes.

Un segundo objetivo, íntimamente ligado con el primero, será poder entender cuál es el efecto específico de cada uno de los factores reseñados sobre la efectividad del mensaje publicado por una organización, de tal manera que se puedan extraer conclusiones prácticas sobre cuáles son las “palancas” que más mueven a los usuarios. Así, a diferencia de los estudios previos, que suelen centrarse en un número reducido de variables, la propuesta realizada en este trabajo adopta un enfoque integral con el que se pretende determinar el efecto relativo de distintas variables relacionadas con los atributos del mensaje y del emisor.

Como contexto de estudio, el trabajo se centra en la Universidad de Zaragoza (UZ), la cual constituye una de las principales universidades públicas españolas. En concreto, para estimar el modelo propuesto se analizarán 53.783 tuits publicados por la UZ tanto en su cuenta principal como en otras 54 cuentas vinculadas a la institución. En las organizaciones sin ánimo de lucro, como son las universidades públicas, no suelen existir partidas presupuestarias para la realización de campañas publicitarias convencionales, por lo que la actividad en redes sociales como Twitter resulta especialmente importante para difundir información, reforzar la reputación institucional y establecer un *engagement* o compromiso con sus públicos (Kuzma et al., 2016; Vogler, 2020).

En definitiva, redes como Twitter constituyen medios muy potentes para conseguir visibilidad y reforzar la reputación corporativa (Schlagwein y Hu, 2016), de modo que la reputación online de una marca puede verse influida por la calidad de sus tuits (Daga et al., 2020). Esta premisa se cumple tanto para las cuentas de marcas de productos como para las marcas corporativas, por lo que cualquier tipo de organización debe velar por la adecuación de sus comunicaciones en estas plataformas. Dentro de la extensa literatura existente sobre Twitter (Karami et al., 2020; Noor et al., 2020), este trabajo pretende realizar una síntesis de los elementos más importantes que pueden determinar el éxito en la difusión de un tuit. De este modo, el objetivo último del trabajo es poder extraer recomendaciones de gestión para las instituciones universitarias en general, y para la UZ en particular, sobre cómo mejorar la efectividad de sus comunicaciones en las redes sociales. Si bien el análisis se centrará en Twitter, los resultados obtenidos podrían ser aplicables a otras redes.

Además de este epígrafe introductorio, el trabajo contempla cuatro apartados en los que se hará un recorrido por el marco teórico, la metodología utilizada, resultados obtenidos y conclusiones generales. Dentro del marco teórico (Capítulo 2), hablaremos sobre la naturaleza de la red social Twitter (Sección 2.1), la investigación académica sobre esta plataforma (Sección 2.2), y la propuesta del modelo y de sus hipótesis subyacentes (Sección 2.3). En la parte de metodología (Capítulo 3) explicaremos brevemente las herramientas utilizadas (Sección 3.1), el contexto objeto de estudio (Sección 3.2), la selección de la muestra (Sección 3.3) y la medición de las variables (Sección 3.4). Dentro del apartado de resultados dedicaremos diferentes subsecciones a la identificación de los elementos de los tuits (Sección 4.1), a su análisis semántico (Secciones 4.2 a 4.4), análisis de legibilidad (Sección 4.5) y estimación final del modelo (Sección 4.6). Por último, en el capítulo sobre conclusiones, implicaciones y trabajos futuros (Capítulo 5) se pretende realizar una reflexión sobre las principales implicaciones académicas y gerenciales del trabajo, así como sobre las limitaciones y recomendaciones de futuras líneas de investigación.

2. Marco teórico

Este capítulo pretende mostrar el estado del arte en cuanto a las características de la red social Twitter y la investigación académica existente. Partiendo de esta base teórica, se realiza una propuesta de un modelo empírico con el que se quiere determinar cuáles son los elementos que más pueden influir en el alcance de un tuit.

2.1. Twitter como red social

Twitter es una notoria red de *microblogging* concebida en 2006 como una plataforma para permitir comunicaciones ágiles entre sus usuarios a través de mensajes cortos. Estos mensajes, denominados como *tuit* en el lenguaje de esta plataforma, tenían una longitud inicial de 140 caracteres. Sin embargo, esta limitación causaba problemas para la comunicación en algunas lenguas, lo que llevó a la compañía a aumentar el número de caracteres hasta los 280 actuales. En palabras de Aliza Rosen e Ikuhiro Ihara, responsable de producto e ingeniero de software respectivamente, “*Intentar comprimir tus pensamientos en un tuit -a todos nos ha pasado-, a veces puede ser complicado*”. Mientras que 140 caracteres resultaban más que suficientes para un japonés, chino o coreano, los usuarios de lenguas derivadas del latín tenían que repensar muchas veces sus tuits para poder adaptar sus mensajes a esta limitación (Jiménez Cano, 2017).

Lo cierto es que Twitter es una red que ha cultivado todo un lenguaje propio. Además de escribir *tweets*¹ los usuarios pueden descubrir las últimas tendencias o *trending topics*, utilizar *hashtags* o términos que permiten seguir y participar en conversaciones sobre un tema determinado (ej. #USElection2020) y *menciones* para dirigirse a usuarios concretos (ej. @unizar), de los que pueden ser seguidores o *followers*, y realizar *retweets* o “republicaciones” de sus mensajes. Estos mensajes no tienen por qué incluir únicamente texto, y pueden dinamizarse mediante recursos como emojis, imágenes o vídeos, e incluso adoptar el formato de pequeñas encuestas. Y como en otras redes sociales, también se permite interactuar con cualquier mensaje, contestando al emisor, compartiendo sus publicaciones o simplemente dándole al *like*. Al alcance de un *tuit*, cualquier persona tiene la posibilidad de convertirse en *tuitero* y empezar a *tuitear*.

Para entender lo que es Twitter y hacia dónde se dirige esta red social resulta interesante examinar su identidad corporativa. En su página web, la compañía se define a sí misma como “*un servicio abierto que alberga un mundo de personas, perspectivas, ideas e información diversas*”. Su misión o propósito general lo definen como la generación de un “*espacio libre y seguro para hablar*”, y entre los valores o principios en los que quieren cimentar su marca se incluyen la promoción de “conversaciones saludables” (1), ganarse la confianza del público (2), la sencillez y transparencia (3), la responsabilidad social (4) y actuar de manera rápida, generando sentimientos de libertad y diversión (5). Por ejemplo, respecto a la responsabilidad social tienen un programa denominado como “Twitter for Good” que aborda diferentes iniciativas centradas en la educación y seguridad en internet, la libertad de expresión y libertades civiles, la promoción de la igualdad, la conservación y sostenibilidad ambiental y la respuesta a crisis y emergencias. La defensa de estos valores no es una cuestión menor, en cuanto a que Twitter se ha compartido en un campo abonado para las *fake news*, para la viralidad de información sesgada hacia determinados intereses, y para la descalificación, con

¹ La Real Academia Española (RAE) recomienda utilizar el término tuit (y sus derivados tuitear, retuitear, etc.), en vez de tweet. En cualquier caso, ambos términos suelen utilizarse indistintamente en las publicaciones en lengua castellana.

mensajes poco éticos o directamente ilícitos. Estos mensajes muchas veces provienen de personas que ocultan su identidad mediante cuentas falsas que, a través de programas denominados *bots*, generan tuits y responden automáticamente a otros mensajes (Liu, 2019).

Figura 2.1. Imagen de Twitter



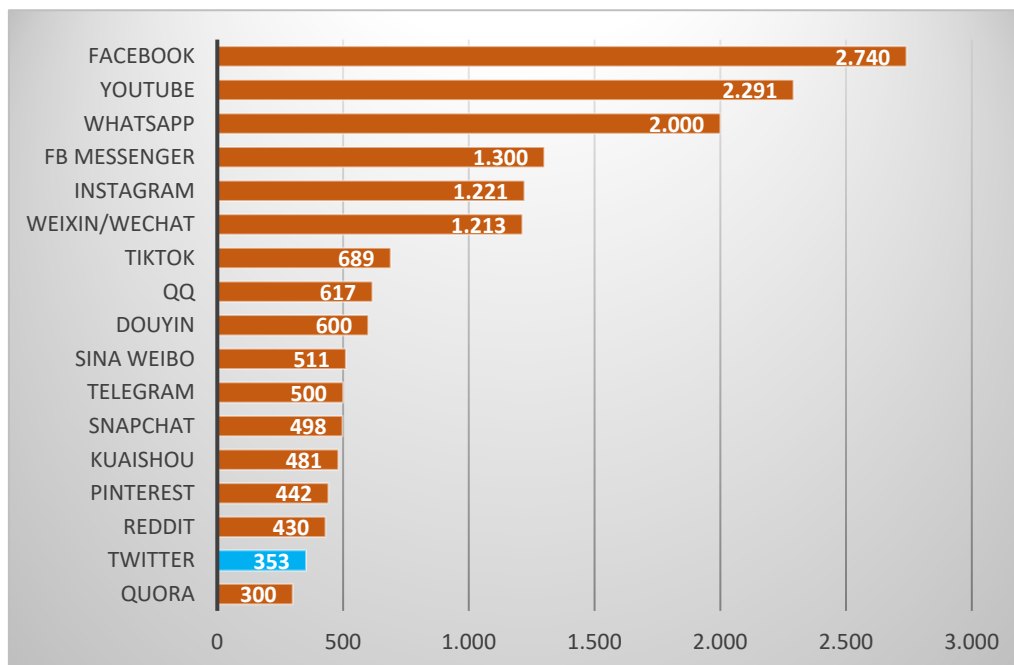
Fuente: <https://twitter.com/explore> (captura tomada el 14/2/2021)

Además de Twitter, dentro de las redes de microblogging han surgido otros actores como [Tumblr](#), aunque con un éxito mucho menor que el de la red del pajarito. Los verdaderos competidores son en realidad otras plataformas que también ocupan el tiempo de los usuarios como [Facebook](#) o [Instagram](#). Aunque el uso de las principales redes globales, Twitter inclusive, está prohibido en China, tampoco podemos pasar por alto el volumen de usuarios que tienen redes como [WeChat](#) y [QQ](#) en el gigante asiático. Y poco a poco avanzan nuevos actores con vocación global como [TikTok](#) y [Twitch](#) (We Are Social & Hootsuite, 2021). Como podemos ver en la Figura 2.2, Twitter se encontraría actualmente en el puesto 16 si solamente nos atenemos al número de usuarios; sin embargo, si descartamos a las redes de carácter “local”, Twitter se alzaría entre las 10 primeras con una comunidad activa superior a los 350 millones de personas.

Con independencia del número de usuarios totales, lo cierto es que en los países occidentales Twitter se encuentra entre las redes sociales favoritas tanto para usuarios como para organizaciones. Por ejemplo, de acuerdo con el estudio realizado por IAB Spain (2020), Twitter sería la tercera red más recordada de manera espontánea en España después de Facebook e Instagram. El mismo estudio también señala que Twitter sigue siendo una de las redes sociales más valoradas por los profesionales, siendo difícil

encontrar una organización que no tenga cuenta activa en Twitter². La capacidad de movilizar a los usuarios a través esta plataforma queda demostrada en contextos como la política.

Figura 2.2. Redes sociales más utilizadas (usuarios globales en enero 2021)



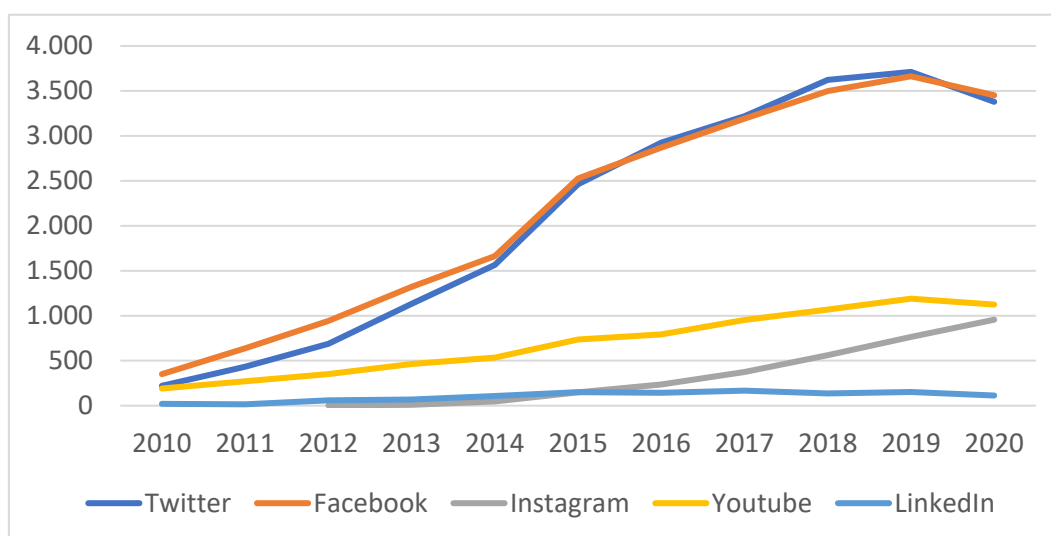
Fuente: We Are Social & Hootsuite (2021)

2.2. Twitter en la investigación académica

En los últimos años, Twitter ha ido adquiriendo un lugar destacado en el mundo académico. Cogiendo como ejemplo a la base de datos Web of Science (WOS), una de las principales referencias para los académicos, podemos ver cómo desde 2010 ha ido creciendo gradualmente el número de investigaciones (artículos, actas de congresos, capítulos de libros, etc.) sobre las principales redes sociales, con excepción de LinkedIn, para la que no se observa un claro crecimiento. En particular, las redes sociales en las que se ha investigado más son Twitter y Facebook, con una evolución muy similar. De hecho, en la Figura 2.3 podemos apreciar como la evolución de las publicaciones en Twitter y Facebook adopta un patrón que se asemeja al ciclo de vida de los productos, apreciándose una cierta fase de introducción (hasta 2012, en el caso de Twitter), crecimiento y tendencia hacia la madurez en este campo.

² Hay que tener en cuenta que este dato recoge tanto cuentas con un volumen de actividad constante, como cuentas prácticamente abandonadas.

Figura 2.3. Evolución del número de publicaciones sobre redes sociales



Fuente: elaboración propia con datos de Web of Science (WOS)

De acuerdo con el estudio bibliométrico realizado por Noor et al. (2020), en el cual se examinaron más de 11.000 publicaciones en WOS sobre Twitter, la mayoría de los estudios realizados con Twitter han sido sobre eventos, tanto de entretenimiento como relativos a catástrofes o crisis, temas centrados en la salud, educación y política. Se ha analizado el rol de Twitter para estudiar las percepciones y emociones de los usuarios, para detectar tendencias, para difundir información y para generar *networking*. En una segunda clasificación, que en este caso se extiende a otras bases de datos, Karami et al. (2020) determinaron que los estudios en Twitter pueden clasificarse en función de su temática, metodología y herramienta de análisis utilizada. En general, el éxito de Twitter como plataforma de investigación se debe a que es un entorno muy rico y dinámico que permite disponer de datos sobre temas sociales que pueden ser abordados mediante diferentes técnicas y procedimientos (Zimmer y Proferes, 2014; McCormick et al., 2017; Stewart y Quan-Haase, 2017).

Entrevistando a una muestra de directivos de grandes compañías, Schlagwein y Hu (2016) identificaron cuatro razones fundamentales para el uso de las redes sociales (Tabla 2.1). Twitter, específicamente, supone un canal ampliamente utilizado por las empresas para difundir información general e interactuar con los usuarios como parte de su “servicio al cliente”, dialogar con ellos y establecer relaciones duraderas (Schlagwein y Hu, 2016; Wang y Yang, 2020). Evidentemente el potencial de Twitter para generar *engagement* trasciende a las empresas con ánimo de lucro, siendo también muy valorado para la consecución de objetivos de carácter social (Gálvez-Rodríguez et al., 2016; Campbell y Lambright, 2020).

Tabla 2.1. Uso de las redes sociales en las organizaciones

Tipos de uso	Características
Difusión: el uso de las redes sociales para la difusión unidireccional de información.	Proporcionar información propia específica Mostrar la propia experiencia profesional Difundir noticias de valor Presentar información de valor Comercializar productos o servicios específicos Promover la marca o la organización en general
Diálogo: el uso de las redes sociales para el diálogo y la comunicación multidireccionales.	Conseguir el compromiso de diferentes áreas funcionales/profesionales Comunicarse con expertos en la materia Adquirir conocimientos de manera fortuita Recibir comentarios sobre un producto o marca Detectar y debatir nuevas ideas
Colaboración: el uso de las redes sociales para facilitar la creación de una solución, producto o servicio en particular.	Crear una nueva solución, producto o servicio Apoyar a un grupo de usuarios Apoyar funciones comerciales específicas Apoyar funciones comerciales generales Gestionar proyectos Colaborar con clientes
Gestión del conocimiento: el uso de las redes sociales para el almacenamiento, mantenimiento y recuperación de conocimiento.	Mantener el conocimiento organizacional Proporcionar un repositorio central Formalizar el conocimiento Mantener un canal de auditoría de la información Almacenar información comercial Permitir la búsqueda de datos
Sociabilidad: el uso de las redes sociales para generar amistad, lealtad y relaciones sociales.	Fomentar el compañerismo dentro de la organización Construir relaciones sociales Conectar grupos con intereses privados similares Facilitar conversaciones no comerciales

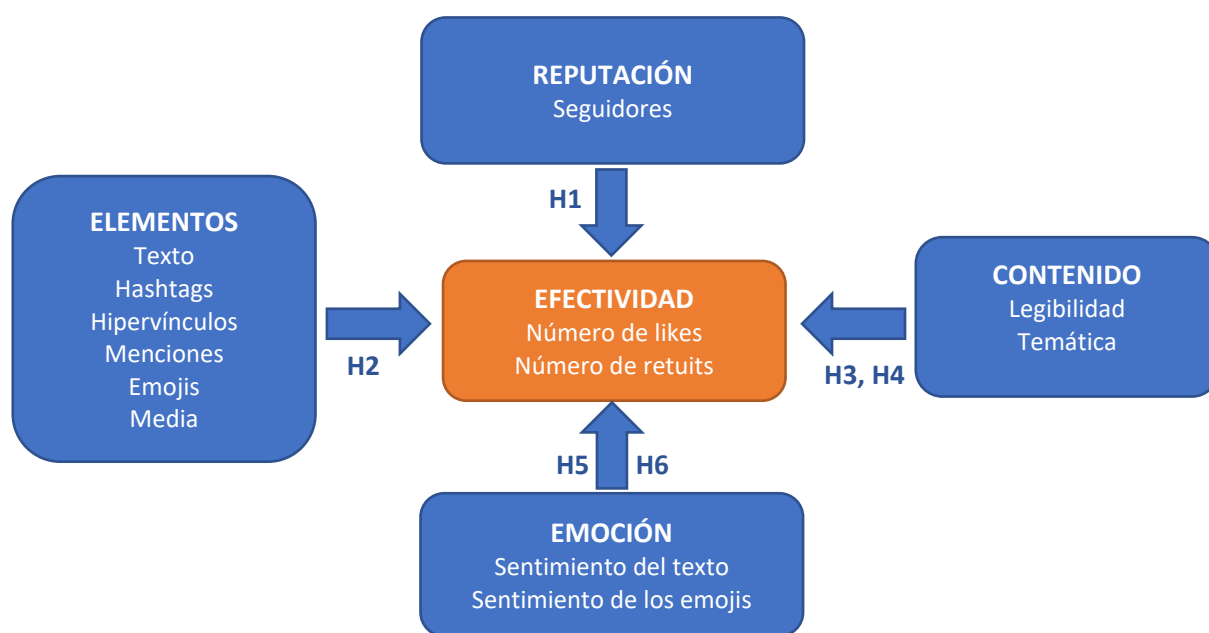
Fuente: Schlagwein y Hu (2016)

2.3. Efectividad de las comunicaciones en Twitter. Modelo propuesto.

Twitter es una herramienta fundamental para la comunicación corporativa, debiendo concebirse como un canal que permite crear marca y conversar con los seguidores. Es importante hacer un uso racional del número y enfoque de los tuits publicados, para conseguir impactar en el público objetivo en el sentido deseado por la organización (Wilcox et al., 2012).

En la Figura 2.4 se propone un modelo sobre la efectividad de las comunicaciones en Twitter, definidas como el número de *likes* o retuits que recibe un tuit publicado en una cuenta determinada. El modelo propuesto recoge algunos factores, aplicados al mundo de Twitter, relacionados con los factores que se consideran primordiales para determinar el éxito de un mensaje en cualquier plataforma, y que básicamente aluden al atractivo del mensaje y al atractivo de la fuente, teniendo en cuenta tanto atributos racionales como emocionales (Wilcox et al., 2012; Kapoor et al., 2021; Wagner et al., 2017; Wajiz et al., 2021). A continuación, describimos cada uno de los elementos incluidos en esta propuesta y las hipótesis objeto de estudio.

Figura 2.4. Modelo sobre la efectividad de las comunicaciones en Twitter



Fuente: elaboración propia

Reputación

La reputación es un factor fundamental que requiere una gestión activa por parte de los responsables de la comunicación y las relaciones públicas (Wilcox et al., 2012). De acuerdo con Fombrun (1996), la reputación constituye una “*representación perceptual de las acciones pasadas y esperables en el futuro de una compañía que describen el atractivo global de la empresa para todos sus constituyentes clave en comparación a otros rivales destacados*”. En definitiva, supone una percepción sobre la competencia de una entidad a lo largo del tiempo, pudiendo utilizarse también este término para valorar el atractivo de cualquier tipo de marca (de Chernatony, 2001).

En la literatura, se ha propuesto utilizar los tuits como indicadores de la reputación de una marca (Liu et al., 2017; Prada e Iglesias, 2020), resultando menos habitual explorar cómo la reputación de la marca es en sí un determinante del éxito de un mensaje en términos de viralidad o *engagement*. En el ámbito universitario, contexto de este trabajo, Vogler (2020) analizó cómo los tuits de las universidades suizas contribuyen a su reputación en función del grupo de interés al que se dirigen, distinguiendo entre académicos, sociedad y medios. Respecto al público principal de las instituciones universitarias, que son sus estudiantes, Kuzma y sus colegas (2016) afirman que estas instituciones “*deben revisar continuamente su estrategia de compromiso social y su reputación online para garantizar que se maximizan todas las oportunidades para mantener y mejorar la inscripción y retención de estudiantes*”.

Dado que la reputación de la fuente determina la credibilidad de un mensaje en Twitter (Westerman et al., 2012), proponemos que la reputación del emisor de un tuit, valorable a través del número de seguidores de la cuenta, puede tener una correlación con la efectividad de dicho tuit. Dicha efectividad se analizará en función de las dos métricas

más importantes que nos proporciona Twitter³, el número de *likes* y el número de *retuits* (Daga et al., 2020). Así, planteamos la siguiente hipótesis de investigación:

H1: La reputación del emisor está asociada con la efectividad de un tuit en términos de likes (a) y retuits (b)

Elementos del mensaje

El siguiente bloque que analizaremos alude a los distintos elementos que podemos encontrar en un tuit como *hashtags*, hipervínculos, emojis y contenidos audiovisuales como fotos y vídeos. La presencia o no de estos elementos podría tener un impacto en la eficacia comunicativa del mensaje.

El hashtag (#), uno de los elementos más populares en el mundo de las redes sociales, contribuye significativamente a la notoriedad de movimientos sociales y tendencias (Wang et al., 2016, Liu et al., 2019). El uso de este símbolo denota que un mensaje puede ser importante dentro de un tema particular y facilita la búsqueda de información sobre dicho tema (Lovejoy et al., 2012). Después de los *hashtags*, los hipervínculos pueden ser los elementos más poderosos para integrar en el tuit contenido de diferentes fuentes y contribuir a una estrategia de comunicación eficaz (Lovejoy et al., 2012; Wang y Zhou, 2015). Estos vínculos pueden llevar a URLs externas con contenidos audiovisuales (ej. un vídeo de YouTube) aunque también en el propio tuit pueden aparecer incrustadas fotos, vídeos y otros recursos. Además, para “dialogar” con otros usuarios tenemos las menciones o alusiones a otras cuentas (@). Todos estos elementos proporcionan una experiencia interactiva al usuario (Swan et al., 2014).

Por supuesto, dentro de los elementos de interés está el texto *per se* del documento, el cual adopta un estilo más informativo que el utilizado en otras plataformas como Facebook (Stone y Can, 2020). Para reforzar la parte emocional del texto no podemos dejarnos a los emojis, que son los caracteres pictográficos que representan expresiones faciales, personas, lugares, o cosas (ej. 😊, 📍). La mayoría de los mensajes online de las personas utilizan emojis, por lo que las organizaciones también han incorporado este lenguaje en redes como Twitter, contribuyendo a que los mensajes encajen más con el público y consigan más *likes* y *retuits* (Li et al., 2019; McShane et al., 2021). En ocasiones, la intensidad emocional de un tuit también se consigue a través del uso de elementos de puntuación como son los signos de exclamación y de interrogación (Teh et al., 2015; Li et al., 2018). En general, proponemos la siguiente hipótesis:

H2: El número de elementos que forman parte de un tuit está asociado con su efectividad en términos de likes (a) y retuits (b)

Contenido

La legibilidad de la información contenida en un mensaje es un elemento clave para su comprensión y posterior difusión. Por ejemplo, la longitud de un tuit ha sido considerada como un indicador indirecto de la dificultad de lectura (Daga et al., 2020; McShane et al., 2021) aunque también se pueden contemplar otras variables relacionadas con la estructura gramatical.

³ Aunque estas métricas puedan considerarse parámetros objetivos para medir el alcance de un mensaje, hay que tener en cuenta el problema que supone la posibilidad de que existan falsos seguidores y la proliferación de *bots* o programas que generan automáticamente *likes* y *retuits* (Liu, 2019).

En la literatura se han propuesto diferentes fórmulas para valorar la legibilidad de un texto, siendo la más conocida la propuesta de Rudolph Flesch (Wilcox et al., 2012). Sin embargo, como podemos ver en la Tabla 2.2, tenemos diferentes alternativas para poder valorar la probabilidad de que un mensaje se entienda.

Tabla 2.2. Fórmulas de legibilidad clásicas

Denominación	Expresión
Flesch	$206,835 - 84,6S/P - 1,015P/F$
Farr-Jenkins-Paterson	$-31,517 + 159,9S1/P - 1,015P/F$
Coke-Rothkopf	$235,87 - 84,44V/P - 1,015P/F$
Coleman	$-37,95 + 116,0S1/P + 148,0F/P$
Dale-Chall	$14,862 - 11,42D/P + 0,0512P/F$
Fog	$3,068 + 9,84S3/P + 0,0877P/F$
Automated Readability Index	$-21,43 + 4,71L/P + 0,50P/F$
Coleman-Liau	$-15,8 + 5,88L/P - 29,59P/F$
Kincaid	$-15,59 + 11,8S/P + 0,39P/F$
P = número total de palabras F = número total de frases L= número total de letras V = número total de vocales S = número total de sílabas S1 = número total de palabras de una sílaba S3 = número total de palabras con tres o más sílabas D = número total de palabras incluidas en la lista Dale Long (lista de 3.000 palabras fáciles)	

Fuente: McCallum y Peterson (1982)

Más que aplicar una fórmula concreta, en este estudio analizaremos cómo la efectividad de un tuit puede depender de tres elementos clave relacionados con la legibilidad: el peso del texto en el tuit (longitud del texto sobre longitud total), el número de caracteres promedio por palabra y el número de sílabas promedio por palabra. Estos dos últimos elementos son contemplados en fórmulas como la de Flesch y ARI (Automated Readability Index). El número promedio de frases no se contempla en el trabajo, dado que los tuits suelen tener muy pocas frases, y su identificación automatizada resulta compleja al utilizarse de manera inadecuada los signos de puntuación y los conectores en muchas ocasiones. Así:

H3: La legibilidad de un tuit está asociada con su efectividad en términos de likes (a) y retuits (b)

Dentro de este bloque relacionado con el contenido del tuit, debemos tener en consideración un aspecto de sentido común. Hay temas que resultan más atractivos que otros para determinados usuarios, lo que debería influir en el comportamiento de dichos usuarios hacia la información. Normalmente, los tuiteros suelen trasladar a Twitter temas de interés que descubren en otros medios de comunicación (Kwak et al., 2010), y también “dialogan” con las marcas sobre sus productos, servicios y promociones (Liu et al., 2017). Algunas interacciones pueden llegar a millones de personas, mientras que otras quedan relegadas a un segundo plano. La siguiente hipótesis recoge esta posibilidad, si bien no se contemplan a priori diferencias en cuanto a temas concretos.

H4: La temática de un tuit está asociado con su efectividad en términos de likes (a) y retuits (b)

Emoción

En el mundo de la publicidad y la comunicación, en general, se asume que la manera de diseñar un mensaje puede ser apelando bien a los beneficios utilitarios, también denominados funcionales, o a los beneficios de carácter emocional. Por ejemplo, un mensaje utilitario podría aludir a la calidad o fiabilidad de un producto, mientras que una argumentación emocional intenta estimular sentimientos positivos, o incluso negativos, que influyan en la actitud y comportamiento de las personas (Zhang et al., 2014; Kapoor et al., 2021). Todavía no hay una respuesta clara a la pregunta de qué tipo de estímulos, racionales o emocionales, son más positivos, pudiendo depender de factores como el contexto o tipo de producto (Wagner et al., 2017).

Con independencia de que un emisor diseñe un mensaje que resalte más un tipo de beneficios u otros, lo cierto es que el comportamiento del consumidor siempre está guiado por un componente actitudinal donde las emociones juegan un papel clave. Esta premisa, defendida décadas atrás por autores como Fishbein y Ajzen (1975), está siendo confirmada por las metodologías más actuales basadas en las neurociencias, las cuales revelan que detrás de cada decisión hay siempre una emoción (Adolphs y Anderson, 2018).

En el caso de Twitter existe interés en analizar el sentimiento asociado a los tuits; por un lado, para valorar la reputación de las marcas y, por otro lado, como estudiaremos en este trabajo, para poder aumentar su eficacia comunicativa (Tellis et al., 2019; Vogler, 2020). Por ejemplo, con una potente muestra formada por 4,1 millones de tuits, Tsugawa y Ohsaki (2017) encontraron que los mensajes negativos tienen una difusión un 25% más rápida que los mensajes positivos y neutrales. Sin embargo, el rol de las emociones en el comportamiento del usuario podría estar supeditado al contexto específico del tuit (Araujo y Kollat, 2018; Tellis et al., 2019).

En este estudio trataremos de diferenciar las emociones generadas por el texto *per se* del mensaje de las emociones que pueden evocar los emojis. Como ya hemos comentado anteriormente, estos pictogramas son ampliamente utilizados en Twitter para generar emociones, siendo un instrumento empleado incluso por cuentas de instituciones oficiales. Además de permitir un lenguaje más cercano con los interlocutores, los emojis pueden complementar, eliminar la ambigüedad y mejorar el significado de los mensajes (Li et al., 2018, 2019). Así, las últimas hipótesis que se proponen son:

H5: El sentimiento generado por el texto de un tuit está asociado con su efectividad en términos de likes (a) y retuits (b)

H6: El sentimiento generado por los emojis de un tuit está asociado con su efectividad en términos de likes (a) y retuits (b)

3. Metodología

Para evaluar el modelo propuesto se han tomado una serie de decisiones que básicamente implican la definición del contexto de estudio, el procedimiento de selección de la muestra y la medición de las variables. Para facilitar al lector la comprensión del trabajo, en primer lugar, explicaremos las principales herramientas metodológicas utilizadas. En el Anexo I también se incluye un breve diccionario sobre los principales conceptos.

3.1. Herramientas utilizadas

Tanto en la descarga de los tuits, como en los análisis de sentimiento, la principal herramienta que se emplea es [Python](#), un lenguaje de programación utilizado en estudios similares y que permite trabajar con volúmenes elevados de datos (Daga et al., 2020). En concreto, se instaló Python 3.9, utilizando la librería [tweepy](#), la interfaz OAuth y códigos diseñados *ad hoc* para los diferentes análisis. Esta utilidad funciona accediendo a la API de Twitter, la cual permite conectar con sus servidores y descargar la información requerida dentro de las condiciones que impone la propia empresa. Para poder utilizar esta API hay que identificarse mediante cuatro códigos individualizados (consumer key, consumer secret, access token, access token secret) que se obtienen al darse de alta en la cuenta de desarrolladores de Twitter.

La riqueza de Python reside en las bibliotecas, también llamadas librerías, que los usuarios han desarrollado para automatizar determinados tipos de tareas y análisis. En particular, en el trabajo se recurre a las librerías [VADER](#) (Valence Aware Dictionary for sEntiment Reasoning) y [TextBlob](#), ampliamente utilizadas para el procesamiento del lenguaje natural (NPL) en el análisis de fragmentos de texto (Al-Natour y Turetken, 2020).

Además de Python, en el trabajo se emplea para algunos análisis el lenguaje de macros [Visual Basic for Applications](#) (VBA). Básicamente, VBA es un lenguaje de programación orientado a eventos que permite ampliar las funcionalidades ofrecidas en los programas de Office. En concreto, se diseñaron algunas funciones de procesamiento de texto en la aplicación de Excel de la versión Microsoft 365.

Para realizar algunas tareas de procesamiento de textos también se utiliza [KNIME](#) (Konstanz Information Miner), en su versión 4.3.0, un paquete de software libre utilizado en minería de datos (Guerra-Velasco, 2008). Esta plataforma funciona creando flujos de datos que visualmente se representan como nodos conectados entre sí. Normalmente se crea un primer nodo, que representa el input o entrada de datos, un último nodo que representa el output o salida, y una serie de nodos intermedios en los que se van definiendo los procesos a realizar.

Por último, en la estimación del modelo propuesto, así como en otros análisis estadísticos, se utiliza el programa [IBM SPSS Statistics 22](#). Este paquete estadístico se usa frecuentemente en Investigación de Mercados y tiene una interfaz amigable y fácil de usar para la realización de todo tipo de análisis univariantes y multivariantes.

3.2. Contexto de estudio

Para evaluar el modelo propuesto se ha elegido la UZ, la cual es una de las 50 universidades públicas que existen en España, y que cuenta con una trayectoria dilatada que se remonta a 1542, año en el que el emperador Carlos V le concedió la categoría de “universidad”. Tal y como se especifica en su misión y visión, la UZ es una institución que se define a sí misma como defensora de la sociedad del conocimiento, constituyendo un motor en la formación e investigación de calidad en Aragón.

De acuerdo con las cifras recopiladas por la propia institución, la comunidad universitaria de la UZ está conformada por unos 40.000 miembros, de los que el grueso corresponde a estudiantes (29.674 estudiantes de grado y máster en curso 2020/21). En el ámbito internacional la institución avanza poco a poco en competitividad, situándose

entre la posición 10 (QS World University Rankings) y 18 de las universidades españolas (Time Higher Education World University Rankings) en 2020.

La comunicación tanto a nivel interno como externo es fundamental para garantizar la competitividad de la institución. La UZ no está sola en su entorno, ya que compite directamente con la Universidad de San Jorge y se enfrenta a otros competidores actuales y potenciales que también ofrecen productos formativos con metodologías offline y online (ej. grados online de la Universidad Internacional de la Rioja). Por otra parte, las cifras nos dicen que, aproximadamente, 2 de cada 3 alumnos procede de otras comunidades, destacando Navarra, la Rioja y Castilla y León⁴. De este modo, para contactar con este público objetivo, y con el resto de sus *stakeholders*, la UZ cuenta con la ayuda de una unidad específica dependiente del Gabinete del Rector: el “Gabinete de Imagen y Comunicación”.

En el documento denominado “carta de servicios”, el Gabinete de Comunicación de la UZ reconoce expresamente que su misión es gestionar la identidad corporativa de la UZ, aspirando a ser el referente de la universidad tanto en comunicación interna como externa. En dicho documento, este organismo también asume una serie de compromisos de calidad con distintos grupos de interés (Anexo II). Esta información identifica como público objetivo a los miembros de la comunidad universitaria, unidades de la Universidad, medios de comunicación y sociedad en general. Por otra parte, en cuanto a las funciones generales de la unidad, podemos incluir las siguientes:

- Elaboración y difusión de información institucional (boletín iUnizar, manual de identidad corporativa, manual de bienvenida, etc.).
- Relaciones con los medios de comunicación (notas de prensa, etc.).
- Asesoría a los miembros de la comunidad universitaria en comunicación.
- Inserción de publicidad e información (publicidad en medios, publicidad exterior, etc.).
- Definición y desarrollo de la estrategia online general y de redes sociales.
- Custodia y actualización del archivo fotográfico de la institución.

Como consecuencia de la relevancia creciente de las redes sociales en la difusión de información, el Gabinete de Imagen y Comunicación creó en julio de 2020 el denominado “Mapa de Redes Sociales”. Básicamente, esta herramienta especifica toda la presencia en redes sociales, tanto de la UZ a nivel institucional, como a nivel de cualquier estructura universitaria (facultades, departamentos, titulaciones, etc.). A fecha de su creación, se contabilizaron 300 perfiles de 22 estructuras universitarias, siendo las redes utilizadas Facebook, Twitter, Instagram, YouTube y LinkedIn. Este mapa es considerado por el propio Gabinete de Comunicación como una parte importante de la estrategia de comunicación de la UZ.

⁴ Para consultar estas y otras estadísticas, la UZ cuenta con un repositorio institucional de datos abiertos (<https://zaguan.unizar.es/collection/opendata?ln=es>).

Figura 3.1. Mapa de redes sociales de la UZ

Sigue a la **Universidad de Zaragoza** en







CATEGORÍA: NOMBRE:

Filtrar

Campus

Campus de Huesca				
Campus de La Almunia				 
Campus de Teruel				
Universidad de la Experiencia				

Centros

Centro Universitario de la Defensa (CUD)				 
Centro Universitario de Lenguas Modernas (CULM)				

Fuente: <https://www.unizar.es/redes-sociales> (captura tomada el 16/2/2021)

La UZ dispone de cuenta institucional en las “Big 5” (Facebook, Twitter, Instagram, LinkedIn y Youtube). En la Tabla 3.1 se muestra el porcentaje de uso de cada una de estas redes para cada una de las categorías mapeadas por el Gabinete de Comunicación. En general, la red predominante es Twitter, si bien, puede observarse que en categorías gestionadas por los estudiantes (ej. asambleas/delegaciones, iniciativas de estudiantes, etc.) tiene una mayor presencia Instagram.

Con el fin de conocer el alcance de las cuentas vinculadas a la UZ, se ha elaborado un estudio descriptivo recogiendo información sobre parámetros como el número de seguidores o publicaciones. La Tabla 3.2 resume la información obtenida para las distintas categorías utilizando la mediana y el rango intercuartílico, en vez de la media y la desviación estándar, para evitar la distorsión que supone la existencia de pocos datos en algunos casos. Los datos fueron obtenidos a mediados de 2021 y se analizaron con el software IBM SPSS Statistics 22.

En el Anexo III se incluye una relación completa de las cuentas examinadas dentro de cada categoría.

Tabla 3.1. Presencia en redes sociales de organismos vinculados a la UZ

Categoría	N	Twitter	Facebook	Instagram	Youtube	LinkedIn
Asambleas/delegaciones estudiantiles	12	58,3%	66,7%	75%	8,3%	0%
Asignaturas	7	100%	0%	0%	0%	0%
Campus	4	75%	50%	25%	75%	25%
Cátedras	18	88,9%	55,6%	11,1%	55,6%	33,3%
Centros	23	91,3%	82,6%	60,9%	52,2%	30,4%
Colectivos universitarios	27	51,9%	44,4%	74,1%	22,2%	11,1%
Colegios Mayores y residencias de estudiantes	10	60%	100%	60%	40%	10%
Departamentos	9	55,6%	22,2%	22,2%	11,1%	11,1%
Deporte	16	12,5%	37,5%	81,3%	6,3%	0%
Egresados	7	0%	0%	0%	0%	100%
Empresas Spinoff y Startup	37	86,5%	81,1%	40,5%	35,1%	54,1%
Iniciativas de estudiantes	7	14,3%	28,6%	71,4%	0%	14,3%
Innovación docente	3	33,3%	33,3%	0%	33,3%	0%
Institutos y grupos invest.	43	90,7%	53,5%	25,6%	34,9%	20,9%
Secciones sindicales	3	100%	66,7%	0%	0%	0%
Seminarios	6	66,7%	33,3%	16,7%	33,3%	0%
Servicios centrales	10	100%	30%	30%	30%	10%
Servicios universitarios	35	62,9%	68,6%	60%	42,9%	8,6%
Titulaciones	17	76,5%	58,8%	41,2%	23,5%	17,6%
Vicedecanatos	2	0%	50%	50%	0%	0%
Vicerrectorados	4	100%	0%	0%	0%	0%
Otros	7	42,9%	42,9%	14,3%	14,3%	14,3%
<i>TOTAL</i>	307	63,8%	48,1%	37,4%	26,9%	20%

Fuente: elaboración propia (datos recogidos del mapa de redes sociales el 19/1/2021)

Como podemos ver en la Tabla 3.2, después de la cuenta institucional (25.200 seguidores), son las cuentas de las secciones sindicales las que tienen más tirón en Twitter (2.263), seguidas por los servicios universitarios (978). En Facebook, la cuenta institucional contaba con 28.189 seguidores en la fecha de estudio, muy a distancia de los campus (3.025) y de los centros (947). Respecto a Instagram, aunque la cuenta general es la que más seguidores tiene (5.173), las cuentas sobre cátedras son las que más publicaciones realizan. En Youtube, la cuenta institucional aparece en el segundo puesto (1.310 suscriptores), detrás de “Otros”⁵ (1.510). Y en el caso de LinkedIn, las cifras dependen mucho de la cuenta examinada, siendo la cuenta institucional la que tiene un mayor número de seguidores (96.270).

En general, las publicaciones de Twitter no suelen tener un elevado número de interacciones, siendo el *engagement* aproximadamente del 0,04% frente al 3% que se consigue en Instagram, la red preferida por la población universitaria⁶. También se observa que, mientras que en el caso de Instagram la cuenta institucional tiene 15 seguidores por cada publicación, en el caso de Twitter la ratio es de 2 seguidores por tuit. Estas métricas nos advierten sobre la necesidad de revisar la estrategia de comunicación en Twitter.

⁵ En “Otros” se incluyen las siguientes cuentas: Diseño de producto y marca, Equipo docente de electromagnetismo unizar, FEUZ (Fundación Empresa Universidad), Geografialia (Revista científica del Departamento de Geografía), Relaciones Internacionales EINA, Relaciones internacionales FECEM, y Tropelías (Área de Teoría de la Literatura y Literatura Comparada de la Universidad de Zaragoza).

⁶ Se ha calculado el número de *likes* entre el número de seguidores sobre una muestra de 20 mensajes.

Tabla 3.2. Métricas de las cuentas de la UZ en redes sociales (mediana y rango intercuartílico)

Categoría	Twitter Seguidores	Twitter Tuits	Facebook Seguidores	Instagram Seguidores	Instagram Public.	Youtube Suscriptor	LinkedIn Seguidores
General (cuenta institucional)	25.200	12.400	28.189	5.173	336	1.310	96.270
Asambleas/del. estudiantes	892 (1.509)	782 (1.849)	355 (1.164)	351 (1.853)	19 (176)		
Asignaturas	176 (906)	1.091 (2.818)					
Campus	836 (1.310)	1.323 (4.243)	3.025 (4.236)	556 (0)	227 (0)	130 (889)	275 (0)
Cátedras	609 (5.601)	1.121 (12.587)	742 (12.625)	2.616 (5.114)	432 (848)	55 (341)	721 (1.665)
Centros	462 (8.406)	454 (5.547)	947 (2.378)	452 (1.184)	86 (660)	56 (3.680)	778 (1.851)
Colectivos universitarios	480 (3.192)	1.466 (22.493)	252 (1.332)	474 (3.316)	84 (355)	151 (203)	16 (76)
Colegios mayores y residencias estudiantes	372 (443)	750 (2.143)	288 (1.610)	313 (538)	84 (257)	49 (148)	24 (0)
Departamentos	658 (953)	284 (2.891)	672 (1.144)	372 (530)	66 (61)	45 (0)	16 (0)
Deporte	621 (206)	2.044 (405)	87 (1.941)	403 (1.496)	13 (1.373)	145 (0)	
Egresados							185 (5.544)
Empresas spinoff y startup	476 (9.931)	495 (14.494)	4.778 (89.873)	3.053 (21.008)	122 (2.003)	62 (5.327)	286 (11.544)
Iniciativas de estudiantes	427 (0)	609 (0)	82 (53)	322 (472)	18 (75)		68 (0)
Innovación docente	240 (0)	1.134 (0)	99 (0)			699 (0)	
Institutos y grupos investigación	245 (3.310)	308 (11.891)	467 (6.115)	300 (580)	64 (339)	59 (751)	460 (11.028)
Secciones sindicales	2.263 (22.265)	415 (385)	224 (233)				
Seminarios	60 (1.362)	19 (1.137)	138 (53)	50 (0)	3 (0)	115 (59)	
Servicios centrales	427 (2.336)	289 (5.339)	544 (960)	569 (118)	71 (20)	10 (184)	45 (0)
Servicios universitarios	978 (6.387)	2.975 (19.800)	877 (6.686)	393 (1.083)	101 (427)	23 (1.080)	2.091 (0)
Titulaciones	409 (3.469)	365 (37.998)	285 (3.675)	259 (781)	84 (525)	19 (49)	117 (916)
Vicedecanatos			254 (0)	165 (0)	59 (0)		
Vicerrectorados	581 (465)	1.429 (2.068)					
Otros	182 (99)	39 (933)	359 (1.431)	351 (0)	135 (0)	1.510 (0)	1.708 (0)

Fuente: elaboración propia (datos recogidos del mapa de redes sociales el 19/1/2021)

Nota 1: en algunas cuentas concretas de Facebook los datos representan a datos de “amigos” en vez de “seguidores”.

En el caso de LinkedIn, algunas cuentas son privadas por lo que los datos representan en este caso a miembros de la plataforma.

Nota 2: la categoría general incluye exclusivamente la cuenta institucional, por lo que los datos corresponden a los totales.

3.3. Selección de la muestra

Los datos recopilados en este estudio corresponden a una muestra de tweets publicados desde 55 cuentas vinculadas a la UZ. En concreto, del mapa de redes sociales descrito en el epígrafe anterior, se seleccionó la cuenta institucional (@unizar), así como las cuentas que representan a los campus, centros, vicerrectorados, servicios centrales y servicios universitarios⁷. El resto de las categorías se descartaron por considerarse bien muy específicas (ej. grupos de investigación, información sindical) o por no corresponder a cuentas “propiedad” de la institución (ej. cuentas de colectivos). En la Tabla 3.3 se indican las cuentas concretas estudiadas, así como los datos referentes a la fecha de creación de la cuenta, el número de seguidores y el número de tuits totales.

La descarga de los tweets se realizó mediante el lenguaje Python, utilizando un código diseñado *ad hoc*⁸ que permitía descargar automáticamente todos los tweets publicados por cualquiera de las 55 cuentas descritas (Anexo IV).

Figura 3.2. Vista de Python durante la descarga de información

```
... from pandas import DataFrame
... outtweets = [[tweet.id_str, tweet.user.name, tweet.user.screen_name, tweet.user.description, tweet.user.locati
on, tweet.user.followers_count, tweet.user.friends_count, tweet.user.favourites_count, tweet.user.statuses_count, tweet
.user.listed_count, tweet.user.default_profile, tweet.user.default_profile_image, tweet.created_at, tweet.favorite_count
, tweet.retweet_count, tweet.full_text.encode("utf-8").decode("utf-8")]
... for idx, tweet in enumerate(all_tweets)]
... df = DataFrame(outtweets, columns=["id", "name", "screen_name", "description", "location", "followers_count", "friends_
count", "favourites_count", "statuses_count", "listed_count", "default_profile", "default_profile_image", "created_at", "fa
vorite_count", "retweet_count", "text"])
... nombre="tuits" + user_name
... filename= "%s.csv" % nombre
... df.to_csv(filename, index=False)
... df.head(3)
...
getting tweets before 1344270945553416193
getting tweets before 1333744390922260479
getting tweets before 1323950488140812287
getting tweets before 1312062675317293055
getting tweets before 1301420247782690815
getting tweets before 1276419969060876288
getting tweets before 1268148062314213377
getting tweets before 1257625677031002112
getting tweets before 1245040663437045759
getting tweets before 1236999364758982655
getting tweets before 1226794748352004096
getting tweets before 1207703182341791743
getting tweets before 1192109547843915775
getting tweets before 1179359138104692735
getting tweets before 1172436767741333507
getting tweets before 1151059522846371841
getting tweets before 1149261496238649343
...
   id          name  ... retweet_count          text
0  1353677085760491520  Universidad Zaragoza  ...           0  https://t.co/BNF28ETvk3
1  1353655964600299520  Universidad Zaragoza  ...           0  https://t.co/ZOYiyWH03w
2  1353641128810213376  Universidad Zaragoza  ...           2  Aprovechamos esta publicación para, a modo de ...

[3 rows x 16 columns]
```

Fuente: elaboración propia

⁷ Las cuentas del Servicio de Informática y Comunicaciones, y el Servicio de Publicaciones fueron descartadas por no haber publicado ningún tuit desde su creación. Se incluye la cuenta “Gestión de cátedras” como servicio de interés general.

⁸ El código utilizado es de elaboración propia, si bien se ha construido integrando fragmentos de código o scripts extraídos de diversas fuentes (Yumi, 2018; Morrissey et al., 2020; Roesslein, 2021). En [GitHub](#) puede accederse a los códigos principales utilizados en el TFM.

Tabla 3.3. Población de cuentas utilizadas en la investigación (1 de 2)

Categoría	Organismo	Cuenta	Creación	Seguidores	Tuits
General	Universidad de Zaragoza	@unizar	15/4/2010	25.200	12.400
Campus	Campus de Huesca	@CampusHuesca	15/10/2011	1671	4484
Campus	Campus de La Almunia	@eupla campus	15/12/2011	836	1323
Campus	Campus de Teruel	@TeruelCampus	15/10/2017	361	241
Centros	Centro Universitario de la Defensa (CUD)	@CUDZaragoza	15/2/2015	1441	1386
Centros	Centro Universitario de Lenguas Modernas (CULM)	@CULMUZ	15/5/2012	500	343
Centros	Escuela de Doctorado	@EscuelaUz	15/11/2018	510	500
Centros	Escuela de Ingeniería y Arquitectura	@EINAunizar	15/9/2012	8487	5560
Centros	Escuela Politécnica Superior de Huesca	@EPSunizar	15/7/2017	717	3181
Centros	Escuela Universitaria Politécnica de Teruel	@eupt unizar	15/12/2018	81	113
Centros	Escuela Universitaria de Turismo de Zaragoza	@etuzgz	15/12/2013	380	828
Centros	Facultad de Ciencias	@Ciencias Unizar	15/1/2016	1078	1684
Centros	Facultad de Ciencias de la Salud y del Deporte de Huesca	@FCSD Unizar	15/5/2015	861	1101
Centros	Facultad de Ciencias Humanas y de la Educación de Huesca	@HuescaFCHE	15/6/2016	163	62
Centros	Facultad de Ciencias Sociales y del Trabajo	@movsociz	15/1/2013	488	1161
Centros	Facultad de Ciencias Sociales y Humanas de Teruel	@fesh teruel	15/9/2019	108	113
Centros	Facultad de Derecho	@Derecho Unizar	15/7/2014	436	326
Centros	Facultad de Economía y Empresa	@FECem unizar	15/11/2017	656	1443
Centros	Facultad de Educación	@educacionunizar	15/12/2012	157	14
Centros	Facultad de Empresa y Gestión Pública de Huesca	@FEGPHuesca	15/2/2018	151	408
Centros	Facultad de Filosofía y Letras	@FacultadFiloZgz	15/7/2020	187	82
Centros	Facultad de Medicina	@MedUnizar	15/2/2020	133	129
Centros	Facultad de Veterinaria	@FVeterinariaUZ	15/6/2015	1379	3369
Centros	Instituto de Ciencias de la Educación (ICE)	@ICE UniZar	15/12/2018	136	13
Vicerrectorados	Vicerrectorado de Estudiantes y Empleo	@VrestUnizar	15/3/2020	348	395
Vicerrectorados	Vicerrectorado de Tecnologías de la Información y de la Comunicación	@vrTICunizar	15/4/2016	813	2463
Servicios universitarios	Área de Cultura	@culturauz	15/9/2011	2943	3169
Servicios universitarios	Área de Proyección Social e Igualdad	@areasocialuz	15/12/2011	974	3199
Servicios universitarios	Biblioteca de la Universidad de Zaragoza	@bibliounizar	15/2/2011	6392	19.800
Servicios universitarios	Biblioteca Hypatia de Alejandría	@Bhypatiaunizar	15/3/2010	1750	7.540
Servicios universitarios	Casa del Estudiante	@casaest unizar	15/3/2011	3033	10.100
Servicios universitarios	Consulta de Endocrinología para animales de compañía (Veterinaria)	@EndocrinoUnizar	15/7/2015	568	1.711

Fuente: elaboración propia (datos de Twitter)

Tabla 3.3. Población de cuentas utilizadas en la investigación (2 de 2)

Categoría	Organismo	Cuenta	Creación	Seguidores	Tuits
Servicios universitarios	Cursos de Español como Lengua Extranjera	@ele_unizar	15/2/2013	419	416
Servicios universitarios	Cursos extraordinarios de verano	@CexUnizar	15/6/2014	448	900
Servicios universitarios	Laboratorio de Microscopías Avanzadas	@LMA_UNIZAR	15/4/2015	235	379
Servicios universitarios	Museo de Ciencias Naturales	@museonat	15/12/2014	1661	4.399
Servicios universitarios	Observatorio de Igualdad	@igualdadunizar	15/3/2018	362	1.117
Servicios universitarios	Oficina de Atención a la Diversidad (OUAD)	@ouad_unizar	15/2/2017	824	3.410
Servicios universitarios	Oficina de Software Libre (OSLUZ)	@Osluz_unizar	15/11/2011	761	5.517
Servicios universitarios	Oficina Verde	@OficinaVerdeUZ	15/4/2010	1533	4.829
Servicios universitarios	Prensas Universitarias de Zaragoza (PUZ)	@PrensasUnizar	15/9/2011	982	2.843
Servicios universitarios	Radio Unizar	@RadioUnizar	15/12/2012	2012	3.166
Servicios universitarios	Servicio de Actividades Deportivas	@deportes_unizar	15/3/2011	1419	1.554
Servicios universitarios	Servicio General de Apoyo a la Investigación - SAI	@SaiUnizar	15/2/2020	177	245
Servicios universitarios	Universa	@UniversaUnizar	15/3/2012	3106	3.106
Servicios universitarios	Universidad de Verano de Teruel	@uvteruel	15/12/2012	1112	1.928
Servicios universitarios	Gestión de cátedras	@CatedrasUnizar	15/10/2012	1355	7947
Servicios centrales	Administración Electrónica	@tramita_unizar	15/4/2015	351	1058
Servicios centrales	Área de Calidad y Mejora	@uzcalidad	15/3/2020	94	94
Servicios centrales	OTRI-Oficina de Transferencia de Resultados de la Investigación	@OTRI_Unizar	15/5/2013	2262	5357
Servicios centrales	Sección de Archivo	@archivounizar	15/6/2018	503	261
Servicios centrales	Sección de Becas	@becas_unizar	15/7/2012	1746	316
Servicios centrales	Secretaría General	@GeneralUnizar	15/10/2020	43	18
Servicios centrales	Unidad de Cultura Científica	@UccUnizar	15/9/2014	2379	4410
Servicios centrales	Unidad de Seguridad	@Unidadseguridad	15/10/2013	168	57

Fuente: elaboración propia (datos de Twitter)

El código elaborado en Python generó un archivo CSV con los tuits correspondientes a cada una de las 55 cuentas analizadas. Al compartir la misma estructura, la integración de los ficheros fue sencilla y se realizó desde Microsoft Excel a través del menú de obtención de datos (Datos-Obtener Datos). En total, se descargaron 54.678 tuits, de los que se identificaron 410 tweets con texto repetido y 485 tuits encabezados por “RT” o “Retweeted”, lo que significa que son mensajes retuiteados. Después de quitar todos estos tuits, se llegó a una colección depurada formada por 53.783 tuits.

Cronológicamente, los tuits publicados abarcan el periodo comprendido entre el 3 de marzo de 2011 y el 25 de enero de 2021. Sin embargo, Twitter establece límites en la descarga de tuits, por lo que hay algunas cuentas donde se ha podido llegar “más atrás” que en otras en la recogida de datos. Por ejemplo, los tuits de la cuenta institucional (2.743) llegan hasta julio de 2019, mientras que en otras cuentas se han podido recuperar tuits mucho más antiguos. Para evitar sesgos producidos por esta restricción, en la estimación del modelo propuesto se considerará la antigüedad del tuit como variable de control.

En la Tabla 3.4 se recogen los datos de las cuentas analizadas en términos del número total de tuits, fecha del tuit más antiguo (fecha inicial), *likes* y retuits. Nuevamente se recurre a la mediana y al rango intercuartílico para representar la información, dado que hay una gran asimetría en la distribución de los datos. Así, podemos observar que muchas cuentas reciben un número de *likes* de “0” en mediana, debido a que tienen muchas publicaciones y son muy pocas las que han obtenido un *feedback* positivo,

Tabla 3.4. Muestra de cuentas utilizadas en la investigación (1 de 2)

Cuenta	N	Fecha inicial	Likes		Retuits	
			Mediana	Rango int.	Mediana	Rango int.
bibliounizar	3127	26/2/2019	2	93	0	45
Bhyptiaunizar	2919	12/1/2017	0	29	0	9
EPSunizar	2817	16/8/2017	2	37	0	13
unizar	2743	11/7/2019	7	471	2	124
culturauz	2717	8/9/2011	0	33	0	18
areasocialuz	2715	16/12/2011	0	17	0	12
CampusHuesca	2499	8/9/2014	0	47	1	23
PrensasUnizar	2203	28/9/2011	0	54	0	18
Osluz unizar	2026	10/9/2015	0	39	1	27
OTRI Unizar	1953	16/5/2016	1	33	1	18
vrTICunizar	1799	26/4/2016	1	47	0	36
UniversaUnizar	1718	12/3/2012	0	27	0	16
FVeterinariaUZ	1699	9/10/2015	0	96	0	34
casaest unizar	1645	8/10/2015	0	32	0	13
uvteruel	1539	8/12/2012	0	37	0	18
Ciencias Unizar	1460	23/1/2016	1	79	0	36
ouad unizar	1382	24/5/2017	3	49	1	21
UccUnizar	1338	8/6/2016	7	113	4	92
OficinaVerdeUZ	1249	2/3/2017	1	53	0	24
museonat	1156	28/12/2015	4	75	1	33
CUDZaragoza	995	20/2/2015	4	264	2	34
RadioUnizar	984	9/2/2013	0	11	0	12
EINAunizar	967	22/1/2016	2	66	1	81
eupla campus	938	9/12/2011	0	61	0	46
FECEM unizar	909	20/11/2017	2	31	1	10

Fuente: elaboración propia (datos de Twitter)

Tabla 3.4. Muestra de cuentas utilizadas en la investigación (2 de 2)

Cuenta	N	Fecha inicial	Likes		Retuits	
			Mediana	Rango int.	Mediana	Rango int.
deportes unizar	787	3/3/2011	1	25	1	44
EndocrinoUnizar	780	6/7/2015	0	43	0	11
CatedrasUnizar	759	4/6/2015	0	16	1	11
movsociz	651	10/1/2013	1	25	1	8
CexUnizar	628	20/6/2014	1	32	0	15
etuzgz	621	25/6/2014	0	7	0	6
FCSD Unizar	471	29/5/2015	2	32	1	15
tramita unizar	422	15/4/2015	1	25	1	15
FEGPHuesca	361	8/2/2018	2	14	0	10
CULMUZ	322	31/5/2012	0	7	0	6
ele unizar	314	10/3/2013	0	23	0	9
becas unizar	302	5/7/2012	0	17	0	24
EscuelaUz	277	9/11/2018	1	23	1	10
TeruelCampus	239	7/11/2017	2	23	0	11
Derecho Unizar	224	1/7/2014	0	118	0	18
LMA UNIZAR	210	16/4/2015	4	26	1	16
archivounizar	180	14/6/2018	6	103	1	59
VrestUnizar	105	3/3/2020	6	30	3	27
SaiUnizar	103	26/2/2020	2	17	1	7
MedUnizar	100	3/3/2020	1	15	0	6
igualdadunizar	82	10/4/2018	0	23	0	17
fesh teruel	67	19/9/2019	2	14	1	7
eupt unizar	59	14/12/2018	1	8	0	5
Unidadseguridad	55	3/10/2013	0	2	0	4
HuescaFCHE	53	16/9/2016	0	5	0	5
uzcalidad	46	26/3/2020	3	16	1,5	23
FacultadFiloZgz	36	24/7/2010	2	53	1	17
GeneralUnizar	17	16/10/2020	2	15	0	9
ICE UniZar	8	20/12/2018	1	7	2	5
educacionunizar	7	11/12/2012	0	0	1	4

Fuente: elaboración propia (datos de Twitter)

3.4. Medición de las variables

Las variables del modelo fueron medidas a través de las métricas que se indican en la Tabla 3.5. En el caso de las variables dependientes los datos fueron directamente obtenidos de la descarga de los tuits mediante tweepy, utilizando los objetos⁹ *favorite_count*, para obtener el número de *likes*, y *retweet_count*, para el número de retuits. De igual forma, para valorar la reputación de la cuenta se utilizó como métrica principal el número de seguidores de la cuenta, información que puede obtenerse mediante el objeto de usuario *followers_count*.

El resto de variables fueron creadas a partir de un análisis de contenido centrado en el texto de cada tuit, obtenido mediante el atributo *full_text*. En el caso del bloque de elementos, se contabilizó la presencia o no de *hashtags*, hipervínculos, referencias a otras cuentas, vídeos y fotos. En cuanto al contenido, se valoró la legibilidad de cada

⁹ En la página de desarrolladores de Twitter se incluye un diccionario donde se explican todas las variables y parámetros disponibles al utilizar la API.
<https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview>

mensaje conforme a su longitud relativa (sobre la longitud total del tuit), el número promedio de caracteres por palabra y el número promedio de sílabas por palabra. También se categorizó cada mensaje en función de su temática y, por último, se valoró el componente emocional mediante un análisis diferenciado del sentimiento generado por el texto y por los emojis incluidos en el tuit.

Para controlar su influencia en las relaciones propuestas, en el análisis también se considera la antigüedad de la cuenta y la antigüedad del tuit (atributo *created_at*), en términos del número de días desde la creación de la cuenta y la publicación del tuit, respectivamente. Asimismo, también se construyó una variable que representa el alcance, la cual agrupa las cuentas en función de su orientación general o local (cuentas de campus y centros frente al resto de cuentas).

Tabla 3.5. Medición de las variables

Constructo	Variable	Métrica
Efectividad	Likes	Número de likes
	Retuits	Número de retuits
Reputación	Seguidores	Número de seguidores de la cuenta
Elementos	Hashtags	Hay hashtags (Sí/No)
	Hipervínculos	Hay hipervínculos (Sí/No)
	Cuentas externas	Hay menciones a otras cuentas (Sí/No)
	Emojis	Hay emojis (Sí/No)
	Multimedia	Hay vídeos (Sí/No), fotos (Sí/No)
Contenido	Temática	Temática del tuit
	Legibilidad	Longitud del texto (sobre longitud total del tuit), número de caracteres por palabra, número de sílabas por palabra
Emoción	Sentimiento emojis	Sentimiento de los emojis del tuit (positivo, negativo, neutro)
	Sentimiento texto	Sentimiento del texto del tuit (positivo, negativo, neutro)
Otros	Antigüedad cuenta	Número de días desde la creación de la cuenta
	Antigüedad tuit	Número de días desde la publicación del tuit
	Alcance	Temática de la cuenta (general, local)

Fuente: elaboración propia

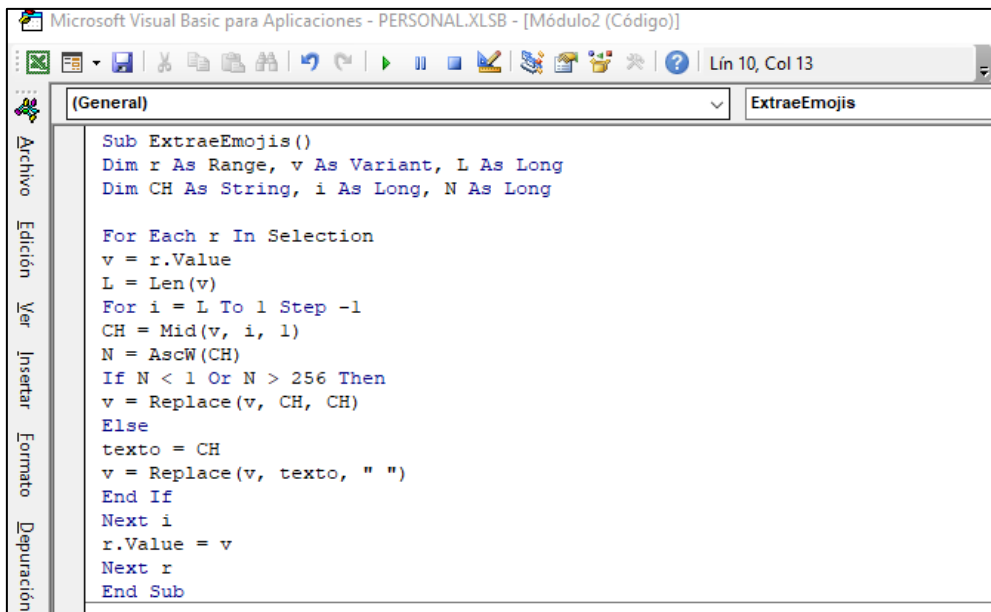
4. Resultados

Para generar las variables descritas anteriormente se utilizó una combinación de técnicas y lenguajes de programación. En los siguientes apartados explicaremos los procedimientos concretos utilizados para identificar cada uno de los elementos objeto de estudio, así como los análisis semánticos con los que se categorizaron los tuits en función de su temática y sentimiento. También mostraremos los criterios utilizados para valorar la legibilidad del texto, cerrando el apartado con la estimación del modelo propuesto en el TFM.

4.1. Identificación de los elementos del tuit

Dentro de un tuit pueden encontrarse diferentes elementos. En primer lugar, para poder identificar y aislar los emojis en una única variable, se utilizó el código VBA mostrado en la Figura 4.1¹⁰. Este código examina mediante la función AscW cada uno de los caracteres del texto. La función devuelve un carácter Unicode, de tal modo que si identifica dicho carácter como “extraño” ($N < 1$ o $N > 256$), la macro lo reemplaza por un espacio. Posteriormente, se revisaron los resultados obtenidos para depurar la información, ya que este código no permite filtrar caracteres como guiones, comillas, símbolos de monedas o caracteres escritos en lenguas como el chino. En total, se identificaron emojis en 6.081 tuits, lo que supone un 11,31% de la muestra.

Figura 4.1. Código para identificar los emojis de los tuits



```
Sub ExtraeEmojis()  
Dim r As Range, v As Variant, L As Long  
Dim CH As String, i As Long, N As Long  
  
For Each r In Selection  
v = r.Value  
L = Len(v)  
For i = L To 1 Step -1  
CH = Mid(v, i, 1)  
N = AscW(CH)  
If N < 1 Or N > 256 Then  
v = Replace(v, CH, CH)  
Else  
texto = CH  
v = Replace(v, texto, " ")  
End If  
Next i  
r.Value = v  
Next r  
End Sub
```

Fuente: elaboración propia

Para la identificación de hipervínculos, *hashtags* y menciones fue suficiente con utilizar la función de texto “hallar”, incluida en Excel. Mediante esta función, se buscaron los caracteres “http”, “#” y “@” dentro de cada tuit, recodificando posteriormente los resultados en tres variables dicotómicas. De acuerdo con los resultados, el 88,28% (47.479) de los tuits publicados por la UZ contienen hipervínculos, el 31,17% *hashtags* (16.761) y el 42,85% (23.043) aluden a otras cuentas. Lógicamente, la inmensa mayoría de los tuits de la muestra también contienen texto elaborado por la propia cuenta (97,47%).

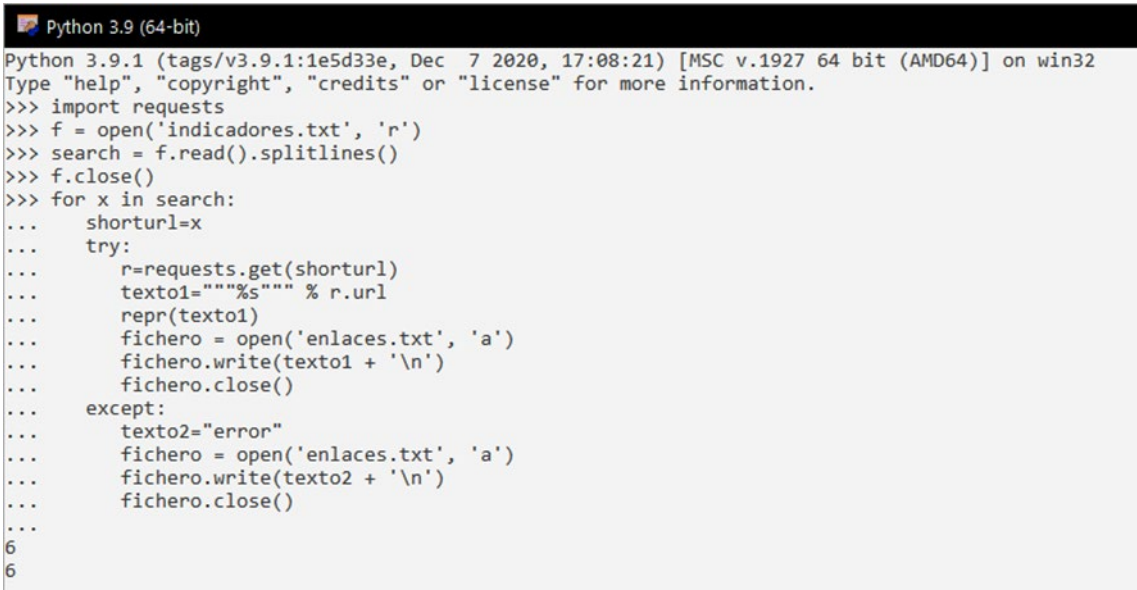
La localización de imágenes y vídeos en los textos de los tuits descargados mediante la API de Twitter es un proceso relativamente complejo. Twitter tiene un objeto denominado *entities* que cuenta con un atributo *media* al que se puede invocar con tweepy, con el fin de descargar los tuits que incluyen elementos como imágenes, vídeos o gifs animados. Sin embargo, este objeto genera problemas al intentar rastrear los tuits con más de una semana de antigüedad, por lo que fue desechado del análisis. Dado que los hipervínculos incluidos en los tuits pueden señalar si un usuario ha adjuntado un

¹⁰ Código adaptado de <https://superuser.com/questions/1543145/remove-emoji-macro>.

elemento audiovisual, la técnica utilizada consistió en analizar dichos hipervínculos de manera automatizada.

Al igual que en otras aplicaciones, los hipervínculos de Twitter se muestran de manera abreviada, de tal manera que por ejemplo la URL “<http://t.co/9lSqhje>”, apunta a una URL extendida¹¹ que comienza por “<https://twitter.com/>” y termina en “/photo/1” para el caso de las fotos, y en “/video/1” para los vídeos. El reto era averiguar cuál era la versión extendida de la URL incluida en el tuit sin tener que “hacer clic” manualmente en cada uno de los enlaces. La solución se halló escribiendo un *script* en Python que partía del listado de las URL abreviadas -incluidas en un fichero de texto-, buscaba una a una las direcciones en Internet, y devolvía las URL extendidas en otro fichero de texto. Es un proceso algo lento, que puede implicar días de ejecución, pero permite detectar errores con mayor facilidad que otros scripts que se diseñaron inicialmente.

Figura 4.2. Código para extraer las URL extendidas de los tuits



```
Python 3.9 (64-bit)
Python 3.9.1 (tags/v3.9.1:1e5d33e, Dec 7 2020, 17:08:21) [MSC v.1927 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import requests
>>> f = open('indicadores.txt', 'r')
>>> search = f.read().splitlines()
>>> f.close()
>>> for x in search:
...     shorturl=x
...     try:
...         r=requests.get(shorturl)
...         texto1="""%s"" % r.url
...         repr(texto1)
...         fichero = open('enlaces.txt', 'a')
...         fichero.write(texto1 + '\n')
...         fichero.close()
...     except:
...         texto2="error"
...         fichero = open('enlaces.txt', 'a')
...         fichero.write(texto2 + '\n')
...         fichero.close()
...
6
6
```

Fuente: elaboración propia

Una vez finalizado este procedimiento, se extrajeron de manera correcta 38.975 URLs, mientras que 8.210 no pudieron ser recuperadas, lo que puede suceder porque el enlace sea incorrecto o porque el recurso haya dejado de estar disponible. Con esta información, y mediante la fórmula de “hallar” de Excel se encontraron 10.320 tuits con el recurso de foto (22,64% de los tuits¹²) y 580 tuits con el recurso vídeo (1,27%). Con el fin de que la muestra de vídeos fuera un poco más elevada, se incluyeron en esta categoría todos los tuits donde la URL apuntaba a Youtube, el canal por excelencia de vídeos. Así, la muestra de vídeos subió al 3,10% (834 vídeos adicionales). Hay que tener en cuenta que en estas métricas no se consideran otros enlaces externos que al incluirse en el tuit puedan generar imágenes o vídeos a la vista del usuario.

¹¹ En este ejemplo: <https://twitter.com/aecaunizar/status/1257386886315868162/photo/1>.

¹² El porcentaje está calculado sobre una total de 45.573 tuits, los 53.783 originales menos los 8.210 cuya URL no pudo ser identificada.

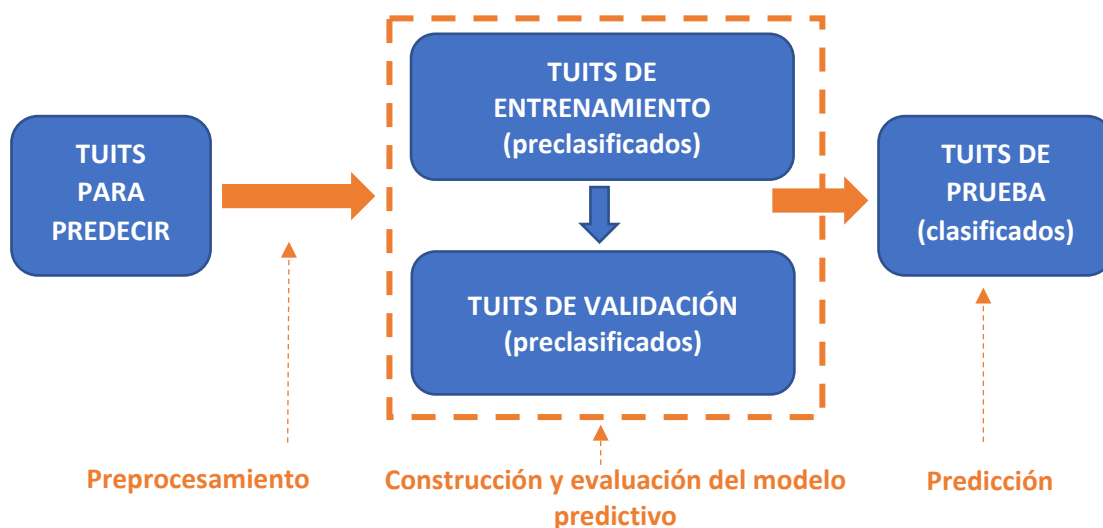
4.2. Análisis semántico I: temática de los tuis

En la última década se han desarrollado herramientas de Inteligencia Artificial que facilitan el procesamiento del lenguaje a través de medios informáticos, término conocido en inglés como *Natural Language Processing* (NLP). Mediante estas herramientas, podemos realizar análisis automatizados del lenguaje, analizando su estructura gramatical, temática, sentimiento, etc. En general, el análisis que se centra en el significado de los textos se conoce en la literatura como “análisis semántico” (Bird et al., 2009; Hardeniya et al., 2016). En este trabajo haremos un primer análisis semántico, centrado en la clasificación temática de los tuis, y un segundo análisis centrado en el sentimiento de los mismos.

El análisis semántico descrito en este apartado tenía el objetivo de categorizar los tuis en función de alguna variable que pudiera ser relevante para la UZ. Dada la importancia que la responsabilidad social tiene en general sobre la reputación corporativa (Wilcox et al., 2012), y que muchas cuentas de la UZ hacen énfasis en este aspecto, se decidió realizar una distinción en función de que los tuis tuvieran o no un componente “social”.

Tal y como mostramos en la Figura 4.3, la metodología concreta utilizada consistió en crear un corpus o conjunto de tuis previamente asignados a las categorías con el que poder entrenar a los algoritmos de clasificación. Durante este entrenamiento, cada tuit se transforma en un vector de palabras y los algoritmos construyen un modelo de clasificación cuya efectividad se prueba sobre otro conjunto de tuis también preclasificados. Una vez que se detecta un algoritmo con una validez predictiva satisfactoria, ya puede utilizarse para clasificar automáticamente cualquier otro conjunto de tuis. Para la realización de todos estos análisis, el texto debe ser previamente preprocesado, estructurando los datos un formato adecuado para los algoritmos.

Figura 4.3. Proceso utilizado en el análisis semántico

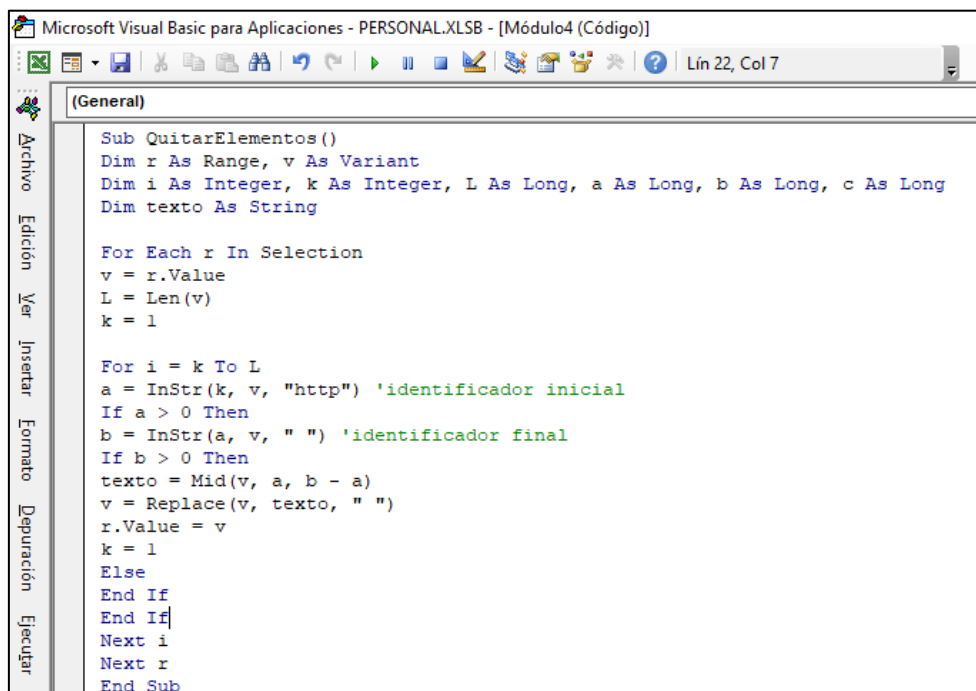


Fuente: elaboración propia

Preprocesamiento de los tuits

Para procesar el texto de un tuit o cualquier otro elemento informativo existen numerosas herramientas. En este trabajo, se optó por realizar una primera depuración de los tuits en Excel, eliminando todos los enlaces mediante una macro¹³ y, posteriormente, utilizando una función personalizada en VBA para eliminar todos los caracteres no alfabéticos (figuras 4.4 y 4.5). Todos los tuits depurados con una longitud inferior a 30 caracteres en la nueva variable de texto fueron descartados del análisis, por no considerarse suficientemente representativos. De la misma manera, se decidió descartar los tuits que empezaban aludiendo a una cuenta externa (@....) por representar una respuesta más que un tuit original. En total, la muestra utilizada de tuits preprocesados alcanza los 47.433 tuits.

Figura 4.4. Código para extraer los *hashtags* del texto



```
Microsoft Visual Basic para Aplicaciones - PERSONAL.XLSB - [Módulo4 (Código)]
Lin 22, Col 7

(Sub)
(General)

Sub QuitarElementos()
Dim r As Range, v As Variant
Dim i As Integer, k As Integer, L As Long, a As Long, b As Long, c As Long
Dim texto As String

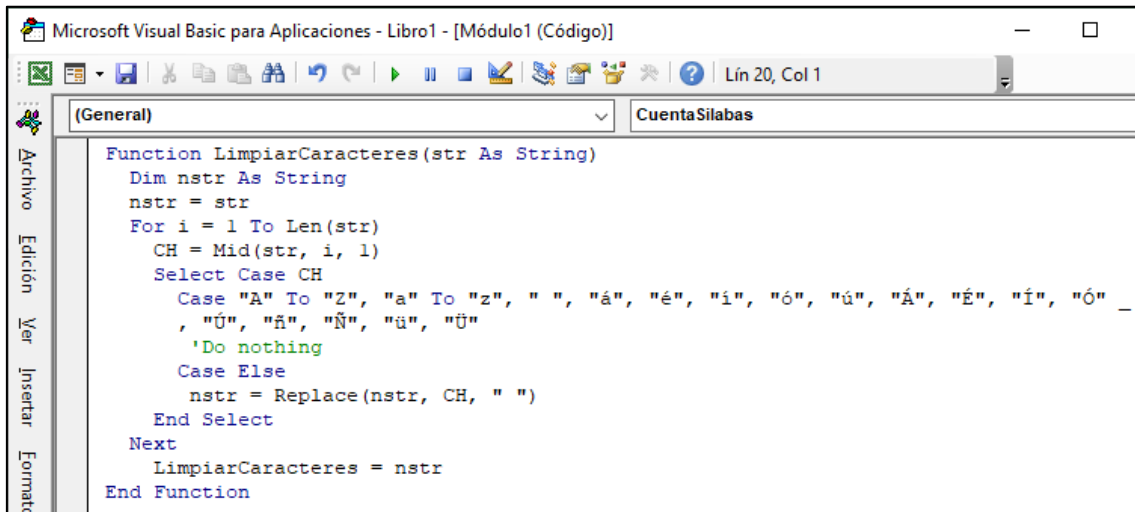
For Each r In Selection
v = r.Value
L = Len(v)
k = 1

For i = k To L
a = InStr(k, v, "http") 'identificador inicial
If a > 0 Then
b = InStr(a, v, " ") 'identificador final
If b > 0 Then
texto = Mid(v, a, b - a)
v = Replace(v, texto, " ")
r.Value = v
k = 1
Else
End If
End If
Next i
Next r
End Sub
```

Fuente: elaboración propia

¹³ Como podemos ver en el ejemplo, el código lo que hace es buscar y eliminar caracteres delimitados por un identificador inicial (http) y un identificador final (un espacio).

Figura 4.5. Código para limpiar los caracteres no alfabéticos

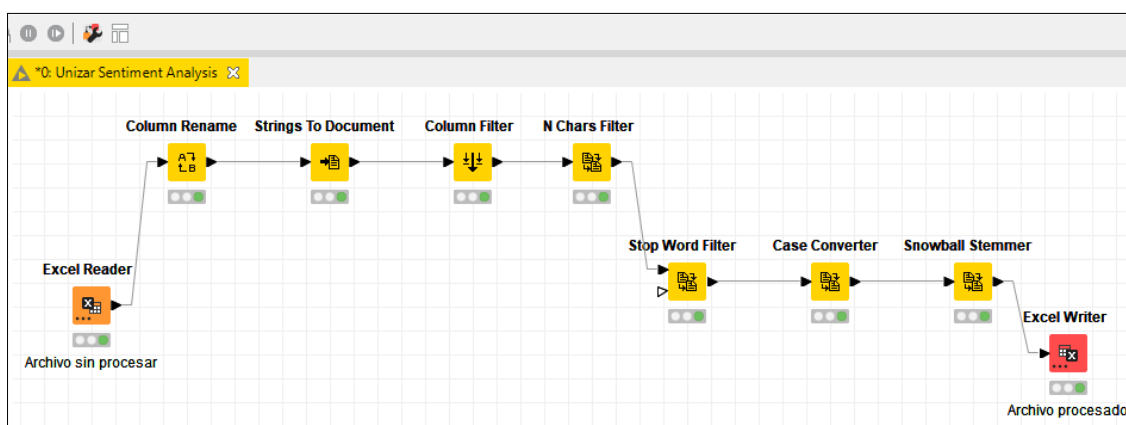


```
Function LimpiarCaracteres(str As String)
    Dim nstr As String
    nstr = str
    For i = 1 To Len(str)
        CH = Mid(str, i, 1)
        Select Case CH
            Case "A" To "Z", "a" To "z", " ", "á", "é", "í", "ó", "ú", "Á", "É", "Í", "Ó", " _
                , "ü", "ñ", "Ñ", "ü", "ü"
                'Do nothing
            Case Else
                nstr = Replace(nstr, CH, " ")
        End Select
    Next
    LimpiarCaracteres = nstr
End Function
```

Fuente: elaboración propia

El texto obtenido fue sometido a un procesamiento adicional con KNIME. Como vemos en la Figura 4.6, el proceso comenzó indicando al programa el fichero XLS con los datos de los tuits en “bruto” (“Excel reader”). El programa realiza algunos ajustes sobre la información del fichero (“columns rename”, “strings to document” y “column filter”), y a continuación empezaría el procesamiento propiamente dicho. Así, lo que se hizo fue eliminar todas las palabras con tres o menos caracteres (“N chars filter”), eliminar las *stopwords* o palabras carentes de significado en castellano (“stop world filter”), convertir el texto en letra minúscula (“case converter”) y aplicar el procedimiento de *stemming* “snowball stemmer”. El *stemming* o radicalización es un procedimiento que consiste en extraer la raíz de las palabras para reducir el número de elementos en el texto y mejorar los resultados de la clasificación. Otro procedimiento alternativo es el de lematización, el cual consiste en transformar algunas palabras por otra común que las represente y que tenga significado pleno. Por ejemplo, cantamos y cantaremos serían transformados a “cant” en un proceso *stemming* y a “cantar” en un proceso de lematización. Aunque a priori, la segunda aproximación pueda parecer mejor, la lematización es mucho más exigente en términos de recursos y se basa en patrones probabilísticos que pueden llevar a transformaciones inesperadas (Urdaneta-Fernández, 2019).

Figura 4.6. Flujo definido en KNIME para procesar el texto



Fuente: elaboración propia basada en KNIME

Creación del corpus del modelo

Para generar un corpus de tuits utilizados con el que entrenar y evaluar nuestro modelo predictivo, se seleccionaron y catalogaron como “sociales” los ítems publicados por las cuentas del Área de Cultura (@culturauz), Proyección Social e Igualdad (@areasocialuz), Museo de Ciencias Naturales (@museonat), Oficina Verde (@OficinaVerdeUZ), Prensas Universitarias de Zaragoza (@PrensasUnizar), Sección de Archivo (@archivounizar), Unidad de Cultura Científica (@UccUnizar), Observatorio de Igualdad (@igualdad), Gestión de cátedras (@CatedrasUnizar), Servicio de Actividades Deportivas (@deportes_unizar), Oficina de Atención a la Diversidad (@ouad_unizar) y Radio Unizar (@RadioUnizar). En total, 12 cuentas con un total de 13.697 tuits.

Dentro del bloque de “no sociales” se seleccionaron 11 cuentas que suelen publicar contenidos sobre servicios e información general para la comunidad de la UZ. En concreto estas cuentas fueron las de Universa (@UniversaUnizar), Sección de Becas (@becas_unizar), Cursos de Español como Lengua Extranjera (@ele_unizar), Oficina de Software Libre (@Osluz_unizar), Cursos extraordinarios de verano (@CexUnizar), Área de Calidad y Mejora (@uzcalidad), Secretaría General (@GeneralUnizar), Unidad de Seguridad (@Unidadseguridad), Servicio General de Apoyo a la Investigación (@SaiUnizar), OTRI-Oficina de Transferencia de Resultados de la Investigación (@OTRI_Unizar) y Administración Electrónica (@tramita_unizar). Estas cuentas suman 6.901 tuits.

Con el fin de que las dos muestras fueran más equilibradas, el bloque de cuentas “no sociales” se reforzó con otros 6.348 tuits que fueron identificados mediante la búsqueda de una serie de palabras clave en las 31 cuentas no asignadas a ninguno de los grupos. Estos términos aluden a la actividad educativa de la UZ y al apoyo a los alumnos en la inserción laboral: cursos, becas, seminarios, formación, workshops, talleres, congresos, conferencias, conferences, symposiums, simposios, jornadas, charlas, aprende, aprender, prácticas, empleo¹⁴.

¹⁴ Se buscaron tanto los términos en plural como en singular, con tildes y sin tildes, y se analizaron muestras aleatorias de tuits para poder detectar y corregir fallos. Por ejemplo, se descartaron los tuits donde el término localizado era “buenas prácticas” o “malas prácticas”, dado que el término deseado era el de prácticas laborales.

Mediante el procedimiento descrito, se eligieron en total 13.697 tuits con una orientación social, y 13.249 tuits con una función más general o utilitaria. Ahora el objetivo era poder construir un algoritmo capaz de predecir si el resto de los tuits encajan en una u otra categoría.

Estimación del modelo de clasificación

La estimación del modelo de predicción se realizó mediante dos algoritmos de clasificación alternativos: Naive Bayes y Logistic Regression (LR). Nuevamente se utilizó Python, utilizando un script donde las palabras preprocesadas dentro de cada tuit fueron convertidas en elementos de una lista, lo que se conoce como *tokenización*, y posteriormente transformadas a números, en un proceso denominado como *vectorización*. Estos números son la materia prima con la que trabajan los algoritmos de clasificación.

En la Figura 4.7 podemos ver un extracto del código utilizado en las estimaciones, pudiendo acceder al código completo en el Anexo V. Como se ha comentado, en el proceso se utilizaron diferentes conjuntos de datos (entrenamiento, validación, prueba), correspondiendo a esta fase la construcción y evaluación del modelo de clasificación mediante los conjuntos que denominaremos “entrenamiento” y “validación”. Para ello, se utilizó una instrucción (`test_size=0.2`) con la que se obliga al algoritmo a usar el 20% de la muestra para evaluar la capacidad predictiva del algoritmo respectivo, utilizando la propiedad (`random_state=0`) para que la “semilla” inicial sea siempre la misma y no se obtengan resultados diferentes en cada estimación¹⁵.

Figura 4.7. Código para construir y evaluar el modelo predictivo (extracto)

```
>>> def load_dataset(filename, cols):
...     dataset = pd.read_csv(filename, encoding = 'latin-1')
...     dataset.columns = cols
...     return dataset
...
>>> def remove_unwanted_cols(dataset, cols):
...     for col in cols:
...         del dataset[col]
...     return dataset
...
>>> def tokenizer(tweet):
...     tweet_tokens=word_tokenize(tweet)
...     tokenized_words=[w for w in tweet_tokens]
...     return " ".join(tokenized_words)
...
>>> def get_feature_vector(train_fit):
...     vector = TfidfVectorizer(sublinear_tf=True)
...     vector.fit(train_fit)
...     return vector
...
>>> dataset = load_dataset("data/clas_modelo.csv", ['ID', 'SOCIAL', 'TEXTO'])
>>> n_dataset = remove_unwanted_cols(dataset, ['ID'])
>>>
```

Fuente: elaboración propia (adaptado de Singhai, 2020)

Para valorar la eficacia predictiva, los algoritmos nos proporcionan diferentes métricas que, cuanto más se aproximan al valor 1, reflejan un mayor éxito en la predicción.

¹⁵ Para minimizar el impacto de la aleatoriedad al dividir los conjuntos de datos, otra opción habitual es utilizar técnicas de validación cruzada. Por ejemplo, mediante una *k-fold cross-validation* de 10 vías ($k=10$), se realizarían 10 procesos de evaluación, dividiendo el conjunto de datos de entrenamiento en 10 particiones diferentes.

Como podemos ver en la Tabla 4.1, los resultados son muy positivos para las dos técnicas de clasificación, sin poder claramente determinar cuál es mejor.

Tabla 4.1. Resultados de los algoritmos de clasificación

Métrica	Cálculo	Naive Bayes	LR
Accuracy (Exactitud)	$(VP+VN) / (VP+FP+VN+FN)$	0,882	0,886*
Precision (Precisión)	$VP/(VP+FP)$	0,901*	0,891
Recall (Sensibilidad)	$VP/(VP+FN)$	0,861	0,882*
F-Measure	$2*(precision*recall) / (precision+recall)$	0,881*	0,886
Matriz de confusión	[VP FN] [FP VN]	[2.412 257] [379 2.342]	[2.375 294] [322 2.399]

Nota: VP: verdaderos positivos, FP: falsos positivos, VN: verdaderos negativos, FN: falsos negativos. VP y VN representan los casos donde el algoritmo acierta en la predicción (Verdad), clasificando como positivo un ejemplo positivo (el tuit es de la categoría social) o negativo un ejemplo negativo (el tuit no es de la categoría social).

Predicción de la categoría de los tuits

La predicción de los tuits no incluidos en el corpus, un total de 20.487, fue realizado tanto con el algoritmo Naive Bayes como el de LR. Para ello se ejecutó un nuevo script en Python (Figura 4.8).

Figura 4.8. Código para predecir la categoría de los tuits

```

Python 3.9 (64-bit)
>>> test_file_name = "data/clas_predice.csv"
>>> test_ds = load_dataset(test_file_name, ['ID', 'CLASE', 'TEXTO'])
>>> test_ds = remove_unwanted_cols(test_ds, ['ID'])
>>> test_ds.text = test_ds["TEXTO"].apply(tokenizer)
>>> test_feature = tf_vector.transform(np.array(test_ds.iloc[:, 1]).ravel())
>>> test_prediction_lr = LR_model.predict(test_feature)
>>> test_result_ds = pd.DataFrame({'CLASE': test_ds.CLASE, 'prediction':test_prediction_lr})
>>> test_result = test_result_ds.groupby(['CLASE']).max().reset_index()
>>> test_result.columns = ['CLASE', 'predictions']
>>> test_result.predictions = test_result['predictions']
>>> fichero = open('predice_lr.txt', 'a')
>>> lista=test_prediction_lr
>>> for item in lista:
...     texto=repr(item)
...     fichero.writelines(texto+ '\n')
...
>>> fichero.close()
>>>
>>> test_prediction_bayes = NB_model.predict(test_feature)
>>> test_result_ds = pd.DataFrame({'CLASE': test_ds.CLASE, 'prediction':test_prediction_lr})
>>> test_result = test_result_ds.groupby(['CLASE']).max().reset_index()
>>> test_result.columns = ['CLASE', 'predictions']
>>> test_result.predictions = test_result['predictions']
>>> fichero = open('predice_bayes.txt', 'a')
>>> lista=test_prediction_bayes
>>> for item in lista:
...     texto=repr(item)
...     fichero.writelines(texto+ '\n')
...
>>> fichero.close()
>>>

```

Fuente: elaboración propia (adaptado de Singhai, 2020)

En el 79% de los casos (16.186 tuits), la predicción de los dos algoritmos fue coincidente. Del 21% restante (4.301), se extrajo una muestra aleatoria de 100 tuits y se revisaron manualmente los resultados para ver qué clasificación tenía más sentido. De una manera evidente, el algoritmo LR mostró patrones más lógicos, lo que confirmaría la asunción de Bhowmik (2015) de que este algoritmo es más potente para discriminar entre categorías. Por tanto, la clasificación realizada por LR será la utilizada en el test del modelo empírico propuesto.

4.3. Análisis semántico II: sentimiento del texto

El análisis del sentimiento de los textos constituye otra metodología de clasificación *per se*, existiendo numerosas herramientas que facilitan la tarea de los investigadores a la hora de predecir si un texto es positivo o negativo. En este trabajo exploraremos tres opciones, la utilización del corpus de TASS, la librería TextBlob y la librería VADER. Previamente, preprocesaremos el texto original de los tuits mediante el procedimiento que se explica a continuación.

Preprocesamiento de los tuits

Para valorar la polaridad del texto de los tuits se construyó una nueva variable en la que además de las URLs también se excluyeron los *hashtags* y las menciones o alusiones a otras cuentas. Además de intentar “simplificar” el tuit, con esta aproximación se quería restringir el sentimiento del tuit al texto realmente “aportado” por el emisor. Para ello, lo que se hizo fue utilizar variaciones de la macro mostrada en la Figura 4.4, cambiando la búsqueda de “http” por “#” y “@”.

Al igual que en el ejemplo anterior, el texto del tuit fue transformado a minúsculas y desprovisto de cualquier carácter que no fuera alfabético mediante el código VBA de la Figura 4.5. Sin embargo, en este caso se decidió mantener la integridad de los tuits, no quitando ninguna palabra y no transformándolos mediante *stemming*¹⁶. En total, se emplean para este análisis 49.701 tuits con una longitud mínima de 30 caracteres.

Corpus TASS

Desde 2021, la Sociedad Española del Procesado de Lenguaje Natural (SEPLN) celebra un Taller de Análisis de Sentimiento en español (TASS). Solicitando previamente autorización, desde su página web se puede acceder a bases de tuits en castellano clasificados de acuerdo con su polaridad y que pueden resultar útiles para entrenar y evaluar los algoritmos (Pla y Hurtado, 2018; Chiruzzo et al., 2020). Específicamente, para ese trabajo se creó un corpus resultado de combinar cuatro bases de datos¹⁷ que alcanzan 15.201 tuits preclasificados en positivos, negativos y neutros¹⁸. Una vez filtrados los tuits por su identificador (ID), se detectaron 6.178 tuits repetidos que fueron descartados para el análisis.

¹⁶ En este tipo de análisis, una palabra como “no” puede cambiar 180 grados la polaridad del tuit, por lo que se ha decidido mantener todas las palabras con independencia de su longitud. Dado que en este análisis se utilizarán librerías consolidadas como VADER, también se ha preferido no modificar el texto mediante procedimientos de lematización o *stemming*.

¹⁷ General-train-tagged-3l: train set (General Corpus, 2012). TASS2018_country_ES_train (Spain dataset, 2018, task1), TASS2018_country_ES_dev (Spain dataset, 2018, task1), TASS2019_country_ES_train: train set (Spain dataset, 2018, task1).

¹⁸ En algunos casos se utilizaban dos denominaciones para los ítems neutros (NEU y NONE) que fueron fusionados en una única categoría.

Para homogeneizar el formato del texto de los tuits con el formato de los tuits a predecir, se aplicaron los procedimientos descritos de eliminar enlaces, *hashtags*, menciones, caracteres no alfabéticos y tuits con poco texto (menos de 30 caracteres), dejando todo el texto en minúscula. Al final, nos quedamos con un corpus TASS de 8.138 tuits, de los que 3.151 tienen una polaridad positiva, 2.827 negativa y 2.160 neutra.

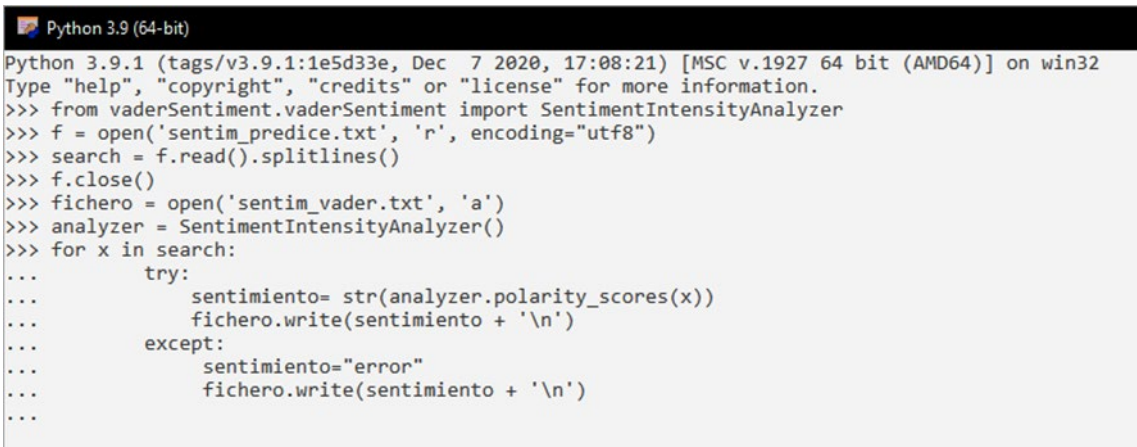
Una vez depurada la base de datos, se ejecutó en Python el mismo código utilizado para construir el modelo de clasificación, descrito en el Anexo V. La base del corpus fue dividida aleatoriamente en un 80% para el entrenamiento del modelo y en un 20% para la evaluación. Sin embargo, en esta ocasión los algoritmos de Naive Bayes y LR no arrojaron buenos valores, con una exactitud (accuracy) de 0,574 para Bayes y de 0,584 para LR. Para ver si estos resultados podían ser mejorados, se realizaron pruebas adicionales ejecutando previamente desde Python procedimientos de *stemming* y de filtrado de palabras. Los resultados obtenidos fueron prácticamente idénticos, por lo que se decidió probar otros métodos alternativos para predecir el sentimiento de los tuits.

VADER

Para el análisis de sentimiento de textos en inglés, probablemente la opción más utilizada en el mundo académico es VADER. Esta librería tiene detrás un potente corpus léxico sobre los sentimientos asociados a los términos utilizados en los tuits (Hutto y Gilbert, 2014), al que algunos autores incluso se refieren como el léxico “de oro” (Al-Natour y Turetken, 2020). Por tanto, este procedimiento puede ejecutarse directamente contra la base total de tuits, sin separar entre conjuntos de entrenamiento, evaluación o predicción.

Para salvar la restricción lingüística simplemente se tradujeron los textos preprocesados del total de tuits (49.701) mediante el servicio de Google Translate. Si bien la traducción se puede también ejecutar mediante código Python, haciendo una llamada a la Google Translate API¹⁹, esta utilidad está restringida a un número de peticiones diarias. Con el procedimiento de traducción indicado, se detectaron 213 tuits que estaban originalmente en inglés, 3 en francés y uno en italiano. Al final, el texto preprocesado, ya traducido al inglés, fue incluido en un fichero de texto (*sentim_predice*) para su análisis por VADER (Figura 4.9).

Figura 4.9. Código de análisis de sentimiento con VADER



```
Python 3.9 (64-bit)
Python 3.9.1 (tags/v3.9.1:1e5d33e, Dec 7 2020, 17:08:21) [MSC v.1927 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
>>> f = open('sentim_predice.txt', 'r', encoding="utf8")
>>> search = f.read().splitlines()
>>> f.close()
>>> fichero = open('sentim_vader.txt', 'a')
>>> analyzer = SentimentIntensityAnalyzer()
>>> for x in search:
...     try:
...         sentimiento= str(analyzer.polarity_scores(x))
...         fichero.write(sentimiento + '\n')
...     except:
...         sentimiento="error"
...         fichero.write(sentimiento + '\n')
... 
```

Fuente: elaboración propia

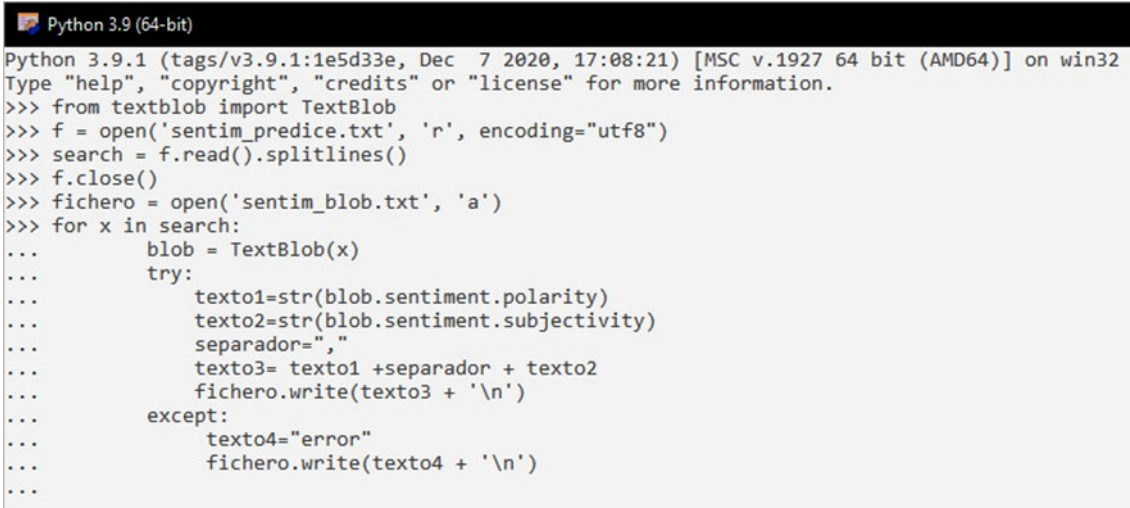
¹⁹ TextBlob, librería que explicaremos en el siguiente apartado, incorpora la opción de traducir un texto a diferentes idiomas, previamente a su clasificación, haciendo una llamada a la API de Google Translate.

La salida de este procedimiento fue otro fichero de texto (*sentim_vader.txt*) donde se recoge el sentimiento asociado conforme a cuatro métricas. Las tres primeras valoran la negatividad, neutralidad y positividad del texto²⁰, oscilando de 0 a 1. La última recoge un *compound* o sentimiento general en una puntuación agregada de -1 a 1. En promedio, la totalidad de tuits de la base de datos tendría un sentimiento positivo (*compound*=0,224) que se confirma por el hecho de que los ítems con polaridad positiva (24.868) son casi 6 veces más que los que tienen polaridad negativa (4.257).

TextBlob

Los investigadores pueden encontrar en *TextBlob* otra alternativa fiable para realizar análisis de sentimiento. De hecho, en algunos contextos se ha encontrado que *TextBlob* podría proporcionar mejores resultados (Al-Natour y Turetken, 2020). A diferencia de *VADER*, *TextBlob* no solamente aplica técnicas heurísticas para analizar el texto, sino que también utiliza una API que le permite realizar tareas de procesamiento de lenguaje natural (PNL) tal como la categorización de información, extracción de frases y el propio análisis de sentimientos. En la Figura 4.10 se muestra el script de Python utilizado sobre la base de 49.701 tuits preprocesados y traducidos.

Figura 4.10. Código de análisis de sentimiento con *TextBlob*



```
Python 3.9 (64-bit)
Python 3.9.1 (tags/v3.9.1:1e5d33e, Dec 7 2020, 17:08:21) [MSC v.1927 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> from textblob import TextBlob
>>> f = open('sentim_predice.txt', 'r', encoding="utf8")
>>> search = f.read().splitlines()
>>> f.close()
>>> fichero = open('sentim_blob.txt', 'a')
>>> for x in search:
...     blob = TextBlob(x)
...     try:
...         texto1=str(blob.sentiment.polarity)
...         texto2=str(blob.sentiment.subjectivity)
...         separador=","
...         texto3= texto1 +separador + texto2
...         fichero.write(texto3 + '\n')
...     except:
...         texto4="error"
...         fichero.write(texto4 + '\n')
... 
```

Fuente: elaboración propia

La salida de *TextBlob* viene dada por dos indicadores. Por un lado, la herramienta intenta valorar la polaridad del texto en un rango de -1 a 1. Por otro, ofrece un valor relacionado con la subjetividad del ítem, que en este caso oscila entre 0 (muy objetivo) y 1 (muy subjetivo). En nuestro conjunto de datos, obtenemos una polaridad ligeramente positiva de 0,123, si bien el número de tuits positivos (22.143) también sobrepasa claramente al de negativos (3.681). La subjetividad promedio es de 0,267 (DT=0,290), por lo que el algoritmo también interpreta que en los tuits tiene más peso la información factual frente a la subjetiva.

²⁰ Tanto en *VADER* como en *TextBlob* el sentimiento del texto es detectado a través de diferentes parámetros donde no solamente se tienen en cuenta las palabras individuales, sino también elementos como la estructura de la frase o la utilización de elementos como signos de exclamación.

Al comparar los resultados con la clasificación realizada por VADER, se encontró que en el 97,31% de los casos cuando TextBlob señalaba un tuit como positivo también lo hacía VADER, sucediendo lo mismo con los ítems negativos. La disonancia entre algoritmos se encontró únicamente en 1.336 tuits, aunque al inspeccionar visualmente la información, se encontraron evidencias claras a favor de elegir la clasificación de VADER para la estimación del modelo final.

Figura 4.11. Comparativa entre los resultados de TextBlob y VADER

	A	B	D	E	F	G	H
1	ID_TUIT	SCREEN_NAME	TEXT	BLOB_SENTIM	VADER_SENTIM	DISAGREE	MAGNITUD
2	49836	unizar	in the con celebrates its participation in the veteran initia	-0,025	0,9382	1	0,9132
3	49526	unizar	congratulations to those curious students involved in the	-0,033333333	0,9255	1	0,89216667
4	32405	igualdadunizar	a small sample of the conference held yesterday morning	-0,075	0,9451	1	0,8701
5	40494	ouad_unizar	if you are pdi or pas de check your email we will forward a	-0,05	0,9109	1	0,8609
6	4183	Bhpatiaunizar	congratulations to yvonne farrell and shelley mcnamara w	-0,1	0,959	1	0,859
7	50633	unizar	they have received extraordinary awards I alumn a jimeno	-0,033333333	0,872	1	0,83866667
8	49844	unizar	a big hug for your commitment to the student body at the	-0,025	0,8625	1	0,8375
9	9839	CampusHuesca	josé antonio adell dynamizes from the microphone the chi	-0,041666667	0,872	1	0,83033333
10	23948	EPSunizar	we congratulate the new rector josé antonio mayoral for h	-0,015909091	0,8442	1	0,82829091
11	50403	unizar	the army posthumously awarded the military commander	-0,016666667	0,8402	1	0,82353333
12	50871	unizar	the counselor enthusiastically congratulates the finalists	-0,1	0,9217	1	0,8217

Fuente: elaboración propia. Los ítems aparecen ordenados en función de la magnitud de la diferencia

4.4. Análisis semántico III: sentimiento de los emojis

El último análisis semántico se dirigió al sentimiento evocado por los emojis incluidos en los tuits. Para ello, se preparó un archivo de texto con los emojis extraídos de los 6.081 tuits en los que se había encontrado este recurso. Dada la elección de VADER para el análisis del sentimiento del texto, directamente se apostó por esta herramienta de análisis. Además, cabe destacar que TextBlob no permite en la actualidad realizar análisis de elementos que no sean de texto.

El código Python utilizado fue el mismo que el mostrado en la Figura 4.9. Los resultados obtenidos revelaron un sentimiento promedio cercano a la neutralidad (0,068), con 982 tuits con emojis que evocan emociones positivas y solo 291 tuits que evocan emociones negativas.

4.5. Análisis de la legibilidad

Determinar el grado de legibilidad de un tuit mediante herramientas automatizadas resulta mucho más complejo que valorar la legibilidad de otro tipo de textos. Esta complejidad surge por el hecho de que el lenguaje que se utiliza en Twitter no es tan estructurado como en otro tipo de comunicaciones, utilizando arbitrariamente los signos de puntuación e intercalando términos que pueden combinar palabras, lo que sucede con muchos *hashtags* (ej. #fotodeldia).

Para facilitar el análisis y su interpretación, en el estudio de legibilidad nos centraremos en el texto “propio” elaborado por el emisor, es decir, en el formato utilizado en el análisis de sentimiento que excluye no solo enlaces, sino también los *hashtags* y menciones.

Empleando Microsoft Excel, se construyó una primera variable con la función LARGO para medir la longitud del texto sobre la longitud total del tuit. Posteriormente, se

crearon las variables número de caracteres por palabra y número de sílabas por palabra. Para obtener el número de caracteres también se empleó la función LARGO; para contar el número de palabras se recurrió a una función que considera los espacios como separadores²¹ y el número de sílabas se obtuvo a través de una función creada en VBA²² (Figura 4.12).

Figura 4.12. Función para contar sílabas

```

Private Function CuentaSilabas(TxtIn As String) As Long
Dim x As Long
Dim l As Long
CuentaSilabas = 0
'Vowels
For x = 1 To Len(TxtIn)
Select Case UCase(Mid(TxtIn, x, 1))
Case "A", "E", "I", "O", "U", "Y"
CuentaSilabas = CuentaSilabas + 1
Case Else
End Select
Next
'Diphthongs
For l = 1 To Len(TxtIn)
Select Case UCase(Mid(TxtIn, l, 2))
Case "IAU", "IAI", "UAI", "UAU", "IEU", "IEI", "UEI", "UEU", "IOU", "IOI", "UOI",
"UOU", "AI", "AU", "EI", "EU", "OI", "OU", "IA", "IE", "IO", "IU", "UA", "UE", "UI", "UO"
CuentaSilabas = CuentaSilabas - 1
Case Else
End Select
Next
End Function

```

Fuente: elaboración propia

Dada que la legibilidad de un texto puede depender significativamente del idioma utilizado, este análisis se aplicó exclusivamente sobre los tuits escritos en castellano, entre aquellos examinados en el análisis de sentimiento. Son un total de 49.484 tuits en los que el texto redactado supone alrededor del 67% del total del tuit. La media de palabras utilizada es 19 (DT=10 palabras), con 2 sílabas por palabra (DT=0,3) y 6 caracteres por palabra (DT=0,8).

4.6. Contraste de hipótesis

Una vez medidas las variables de manera adecuada, la última fase del estudio consistió en contrastar las hipótesis del modelo propuesto (figura 2.4) con el software IBM SPSS. En concreto, con el procedimiento de mínimos cuadrados ordinarios se estimaron dos regresiones múltiples jerárquicas, la primera para la variable dependiente “likes” y el segundo para la variable dependiente “retuits”.

Ambas regresiones jerárquicas comenzaron con la estimación de un primer modelo donde solamente se incluyen las variables independientes *tiempo_cta* (número de días desde la creación de la cuenta) y *tiempo_tuit* (número de días desde la publicación del tuit). De esta manera, se intenta controlar el efecto del paso del tiempo, y aislarlo del resto de variables, tanto en lo referente a la antigüedad del tuit, como en la antigüedad de la propia cuenta. En un segundo paso se añadieron las variables *cta_seguidores*

²¹ LARGO(Xi)-LARGO(SUSTITUIR(Xi;" ";""))+1

²² El código incorpora partes de scripts encontrados en foros como mrexcel.com y openoffice.es

(número de seguidores de la cuenta), *cta_alcance* (alcance de la cuenta; 1: general, 0: campus y centros), *fotos* (hay fotos; 1: sí, 0: no), *videos* (hay vídeos; 1: sí, 0: no), *emojis* (hay emojis; 1: sí, 0: no), *links* (hay hipervínculos; 1: sí, 0: no), *hashtags* (hay *hashtags*; 1: sí, 0: no), *menciones* (hay menciones a otras cuentas; 1: sí, 0: no), *social* (temática del tuit; 1: social, 0: otros), *textopropio* (relación entre el número de caracteres de texto propio del tuit y el total de caracteres; valor de 0 a 1), *caract_palab* (relación entre el número de caracteres y el número de palabras), *silab_palab* (relación entre el número de sílabas y el número de palabras), *texto_sentim* (sentimiento del texto en escala de -1 a 1) y *emojis_sent* (sentimiento de los emojis en escala de -1 a 1).

En la Tabla 4.2 podemos observar que la mayoría de las variables estudiadas ejercen efectos significativos tanto sobre el número de *likes* como sobre el número de tuits. En el caso de los “*likes*”, el modelo completo es significativo ($F=279,725$; $p<0,001$) y supone una mejora, también significativa ($p<0,001$) sobre el modelo que únicamente incluye las variables temporales. Del mismo modo, el modelo que explica los “*retuits*” alcanza significatividad estadística al 99% de confianza ($F=157,138$; $p<0,001$) y supone una mejora incremental sobre las variables temporales ($p<0,001$). Aunque los R^2 ajustados no son muy elevados (0,454 para el modelo de *likes* y 0,318 para el de *retuits*), superan el umbral mínimo de 0,1 que Falk y Miller (1992) fijan para los estudios realizados en el ámbito de las ciencias sociales. Por otra parte, todas las correlaciones bivariadas entre las variables independientes fueron inferiores a 0,7 (Tabachnick y Fidell, 1996), lo que evita el problema de multicolinealidad.

Tabla 4.2. Estimación de los modelos de regresión jerárquica

(N=53.783)		LIKES			RETUITS		
Modelo	Variable	β (Est.)	t-valor	Sig.	β (Est.)	t-valor	Sig.
M1	tiempo cta	-0,096	-5,740	<0,001***	0,018	0,989	0,323
	tiempo tuit	0,633	37,692	<0,001***	0,461	25,696	<0,001***
M2	tiempo cta	-0,080	-4,501	<0,001***	0,057	2,865	0,004***
	tiempo tuit	-0,486	-10,637	<0,001***	-0,457	-8,955	<0,001***
	cta seguidores	0,345	22,769	<0,001***	0,330	19,508	<0,001***
	cta alcance	0,274	9,826	<0,001***	0,248	7,948	<0,001***
	fotos	0,135	10,440	<0,001***	0,068	4,683	<0,001***
	videos	0,046	4,230	<0,001***	0,002	0,161	0,872
	emojis	0,046	0,190	0,849	0,093	0,344	0,731
	links	0,105	2,356	0,019**	0,169	3,416	0,001***
	hashtags	0,070	4,282	<0,001***	0,054	2,952	0,003***
	menciones	0,138	8,905	<0,001***	0,112	6,461	<0,001***
	textopropio	0,012	0,176	0,860	0,018	0,245	0,807
	caract_palab	-0,193	-1,889	0,059*	-0,241	-2,107	0,035**
	silab_palab	0,408	2,284	0,022**	0,312	1,560	0,119
	social	0,051	3,017	0,003***	-0,036	-1,879	0,060*
	texto_sentim	0,041	3,079	0,002***	0,005	0,362	0,718
emojis_sentim	0,038	3,576	<0,001***	0,035	2,955	0,003***	

Nota: * $p<0,1$ ** $p\leq 0,05$ *** $p\leq 0,01$

En primer lugar, cabe destacar que las variables temporales ejercen efectos significativos sobre el número de *likes* y *retuits*. Sin embargo, su efecto no es claro y se

observa incluso un cambio de signo en el coeficiente de la antigüedad del tuit entre el primer y segundo modelo para el caso de la antigüedad del tuit. Estos patrones corroboran la importancia de controlar la variabilidad de estas variables en nuestro modelo para minimizar los sesgos en la estimación de los coeficientes de las variables principales.

Confirmando las hipótesis H1a y H1b, podemos ver cómo un mayor número de seguidores (*cta_seguidores*) se relaciona con una mayor efectividad en las publicaciones de la cuenta, de tal modo que el coeficiente beta estandarizado toma un valor alrededor de 0,3 para ambas variables dependientes ($p < 0,001$)²³. Asimismo, el efecto del alcance de la cuenta (*cta_alcance*) también consigue una magnitud relevante ($\beta_L = 0,274$; $\beta_R = 0,248$) y significativa ($p < 0,001$).

En cuanto a los elementos presentes en el tuit, el único que no se relaciona significativamente ni con los *likes* ni con los retuits es la presencia de emojis ($p > 0,1$). Salvo esta excepción, existe evidencia empírica a favor de las hipótesis H2a y H2b. Así, podemos ver cómo en el caso de los *likes* resulta relevante, y por este orden, incluir menciones ($\beta_L = 0,138$, $p < 0,001$), fotos ($\beta_L = 0,135$, $p < 0,001$), *links* ($\beta_L = 0,105$, $p = 0,019$) y, en menor medida, vídeos ($\beta_L = 0,046$, $p < 0,001$). Por otra parte, para el caso de los retuits, lo más importante parecen ser los *links* ($\beta_R = 0,169$), seguido por las menciones ($\beta_R = 0,112$), ambos significativos al 99% de confianza. Los vídeos no consiguen en este caso estimular el retuit ($p > 0,1$), aunque estos resultados deben tomarse con cautela dado que la muestra de tuits con vídeos identificados es relativamente reducida.

Respecto a la legibilidad de la información, detectamos algunos patrones de interés. Aunque la presencia de texto no influye significativamente en la efectividad del tuit ($p > 0,1$), sí que se detecta una relación significativa negativa entre el número de caracteres por palabra y el número de *likes* ($\beta_L = -0,193$, $p = 0,059$) y retuits ($\beta_R = -0,241$, $p = 0,035$). Asimismo, el número de sílabas por palabra ejerce un claro efecto significativo sobre el número de *likes* ($\beta_L = 0,408$, $p = 0,022$) que no alcanza a ser significativo en el caso de los retuits ($p > 0,1$). Podemos decir que los resultados ofrecen un apoyo moderado a las hipótesis H3a y H3b centradas en la legibilidad del tuit.

La variable social ejerce efectos muy moderados, aunque significativos, para el caso de los *likes* ($\beta_L = 0,051$, $p = 0,003$), y en menor medida para los retuits ($\beta_R = -0,036$, $p = 0,060$). En función de los signos de los coeficientes podemos interpretar que los tuits con un contenido social tienen más probabilidades de obtener un “me gusta”, aunque son otro tipo de contenidos los que más se comparten. Por tanto, tampoco podemos rechazar las hipótesis H4a y H4b que planteaban un efecto de la temática del tuit sobre su efectividad.

Aunque el coeficiente es muy reducido, el sentimiento asociado al texto del tuit tiene un efecto significativo sobre el número de *likes* ($\beta_L = 0,041$; $p = 0,002$), aunque no sobre el número de retuits ($p > 0,1$). Por tanto, los resultados avalan la hipótesis H5a pero no la hipótesis H5b. En línea con las hipótesis H6a y H6b, el sentimiento asociado a los emojis sí que parece ejercer efectos significativos ($p < 0,01$) sobre ambas variables dependientes, aunque con un coeficiente también reducido ($\beta_L = 0,038$; $\beta_R = 0,035$). Por tanto, la presencia *per se* de estos elementos no parece tan importante como el tipo concreto de emoji utilizado.

²³ Si bien, resulta esperable que los tuits publicados por las cuentas con un mayor número de seguidores obtengan más *likes* y retuits, el valor del coeficiente estandarizado (0,3) indica que la difusión de los tuits depende en mayor medida del efecto conjunto de otras variables.

5. Conclusiones, implicaciones y trabajos futuros

En este trabajo hemos tratado de entender qué factores influyen en la efectividad de un tuit, entendido como el número de *likes* y el número de retuits que dicho tuit recibe. Para ello, nos hemos centrado en las principales cuentas gestionadas en Twitter por la Universidad de Zaragoza, una muestra donde se incluye la cuenta corporativa (@unizar) además de otras 54 cuentas de interés general. Los datos han sido obtenidos y analizados mediante una combinación de metodologías y técnicas que incluyen programación en Python, programación en VBA y utilización de herramientas estadísticas y de minería de datos como KNIME e IBM SPSS.

A nivel teórico, el trabajo apunala la importancia de la reputación de la fuente como elemento fundamental de la credibilidad de los tuits (Westerman et al., 2012). Así, las cuentas con más seguidores de partida y que tienen un alcance más amplio (cuenta corporativa, cuentas de servicios, etc.) parecen conseguir más *likes* y retuits. En el ámbito concreto objeto de estudio, los resultados también avalan la importancia de interactuar con el usuario a través de los elementos del tuit (Swan et al., 2014). Elementos audiovisuales, *hashtags*, URLs y menciones a otras cuentas sí que parecen conseguir este efecto y aumentar la efectividad del tuit, aunque en el caso de los emojis los resultados no consiguen corroborar el efecto positivo reportado en otros contextos (Li et al., 2019; McShane et al., 2021).

Implicaciones académicas

El trabajo contribuye a la literatura previa presentando un modelo que permite entender cómo la efectividad de un tuit se construye a partir de distintos factores relacionados con la propia cuenta, los elementos incluidos en dicho tuit, la tipología y legibilidad de los contenidos informativos, y la emoción derivada de la información. En primer lugar, respecto a las emociones, los resultados están en línea con los trabajos que sugieren que el efecto del tipo de polaridad está supeditado al contexto objeto de estudio (Araujo y Kollat, 2018; Tellis et al., 2019). En este contexto educativo, parece importante evitar emociones negativas tanto en los emojis, en el caso de usarlos, como en el texto en general, sobre todo si queremos conseguir *likes*.

Como era esperable, la temática del tuit también aparece como un factor clave en la respuesta del usuario. Se confirma la importancia de la responsabilidad social para conseguir una actitud positiva hacia la marca (Wilcox et al., 2012), aunque se observa que gustar y compartir mediante retuits está supeditado a la temática del tuit. Los tuits con carácter social podrían conseguir más *likes* y los de carácter informativo más retuits. Por otra parte, aunque los tuits son formatos con una extensión muy limitada, los resultados muestran una penalización de los textos prolijos que incluyen muchos caracteres pero pocas sílabas por palabra.

Implicaciones gerenciales

En cuanto a implicaciones prácticas, los resultados obtenidos sugieren una serie de líneas de actuación que, si bien se dirigen a la UZ en particular, podrían ser adoptadas por otro tipo de instituciones similares. En primer lugar, resulta importante aumentar la base de seguidores de las diferentes cuentas, dado que lógicamente este factor ayudará al éxito y difusión de cualquier tuit que se realice. Para ello, la institución y sus entidades asociadas tendrían que realizar un esfuerzo activo, animando a la comunidad universitaria a seguir las redes sociales como un canal en el que pueden informarse e interactuar con dicha comunidad.

A pesar de que estamos en un medio donde el texto es el protagonista, los usuarios responderán mejor a los tuits que llevan elementos audiovisuales como fotografías. En una sociedad que gira en torno a la imagen y los estímulos visuales, debería evitarse la utilización de tuits planos con solo texto. Por otra parte, aunque los *hashtags* son también importantes, los resultados sugieren que el principal elemento para atraer *likes*, conseguir retuits y en última instancia nuevos *followers* es la interacción con otras cuentas a través de menciones.

Los resultados parecen avalar la idea de que la UZ es una institución seria de la que la comunidad espera “cosas serias”. Así, los emojis no parecen ser de especial relevancia para mejorar la efectividad de los tuits, y en el caso de utilizarlos debería evitarse la utilización ya no solo de tuits sino de texto en general que evoque sentimientos negativos. Asimismo, los usuarios parecen valorar positivamente los tuits que cuentan con enlaces a contenidos que les permite aumentar la información sobre un tema concreto.

La legibilidad no es una cuestión que deba pasarse por alto. Si bien nos encontramos ante una comunidad formada, los tuits recargados en los que se suceden términos complejos con muchos caracteres y pocas sílabas parecen ser menos efectivos. Aquí la recomendación sería pensar que no estamos elaborando un artículo académico sino una píldora informativa en la que el emisor y receptor deben compartir el mismo código, un código que no caiga en el vulgarismo, pero que no sea excesivamente formal.

La pregunta más difícil es probablemente la de qué tipo de temáticas debemos abordar en las comunicaciones. Los resultados obtenidos en este trabajo sugieren que las cuentas de los campus y centros no funcionan tan bien como el resto de las cuentas, de tal modo que sería necesario reflexionar sobre qué tipo de contenidos ofrecer en cada caso para generar mayor *engagement* y con qué frecuencia. Una pista nos la da la variable que tiene en cuenta la orientación social, o no, de los tuits. A los tuiteros les gusta encontrar información sobre las actuaciones en materia de responsabilidad social que realiza la UZ pero, en última instancia, la información que deciden compartir en forma de retuit podría versar sobre contenidos de carácter más informativo o que sienten que les implica de manera directa.

Limitaciones y líneas de investigación

El estudio realizado no está exento de limitaciones. Debido a las restricciones impuestas por la propia API de Twitter no ha sido posible aumentar el número de tuits de la muestra o extraer información más precisa sobre la presencia en los tuits de elementos audiovisuales. Las clasificaciones realizadas por los algoritmos para categorizar los tuits y el sentimiento de los mismos tampoco son exactas, existiendo un margen de error que para subsanarse requeriría una revisión manual de la totalidad de tuits. En la estimación del modelo de regresión también hay que considerar que no se dispone del mismo número de observaciones en cada variable independiente, de tal modo que los coeficientes han sido calculados utilizando los casos con información completa para cada una de las variables. Esta aproximación, al igual que otras utilizadas por IBM SPSS para el tratamiento de los datos que faltan (*missing data*), suele llevar a la obtención de estimadores sesgados que reducen la fiabilidad de los resultados (Von Hippel, 2004).

Para seguir avanzando, en futuros estudios sería interesante profundizar en el rol de la temática de los tuits, realizando nuevas categorizaciones bien mediante procedimientos automáticos o humanos. También podría realizarse un análisis más exhaustivo de la legibilidad de la información utilizando librerías especializadas de Python como

Readability. Este análisis podría confirmar si la complejidad lingüística del texto afecta al éxito de un tuit, tal como los resultados obtenidos en la estimación del modelo parecen indicar.

Otra línea de investigación sería analizar cómo la respuesta hacia los tuits depende del público objetivo al que se dirigen (profesores, estudiantes, general, etc.), en la línea del trabajo de Vogler (2020). Asimismo, sería interesante analizar en profundidad los efectos de interacción que pueden producirse entre las distintas cuentas que pertenecen a la misma institución, viendo en qué grado cada una de ellas contribuye a construir la reputación general de la UZ. Tampoco podemos perder de vista las comunicaciones realizadas por otras universidades e instituciones educativas y que, en temas como la oferta de titulaciones, compiten por conseguir la atención e interés del mismo público objetivo.

Ya para terminar, como predecía Qualman (2012), debemos abundar en la idea de que las redes sociales son los motores de una auténtica revolución en el modo de comunicación y relación entre personas, y entre las personas y las organizaciones. La pregunta ya no es si debemos estar ahí, sino cómo estar ahí. Y los protagonistas cambian y se mueven a toda velocidad. Aunque Facebook ostenta la corona de la red social con mayor número de usuarios a nivel mundial, la población joven -universitaria inclusive- está abandonando esta plataforma. Instagram se afianza como la red visual de referencia, mientras que poco a poco se le van acercando rivales como Tik Tok, Y en medio tenemos a Twitter, un gran canal para conversar con nuestro público objetivo, crear reputación y marca.

Referencias

- Adolphs, R. & Anderson, D. J. (2018). *The Neuroscience of Emotion: A New Synthesis*. Princeton: Princeton University Press.
- Al-Natour, S. & Turetken, O. (2020). A comparative assessment of sentiment analysis and star ratings for consumer reviews. *International Journal of Information Management*, 54, 102132.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.
- Bhowmik, T. K. (2015). Naive bayes vs logistic regression: theory, implementation and experimental validation. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 18(56), 14-30.
- Campbell, D. A., & Lambright, K. T. (2020). Terms of engagement: Facebook and Twitter use among nonprofit human service organizations. *Nonprofit Management and Leadership*, 30(4), 545-568.
- Chiruzzo, L., Rosá, A., & Etcheverry, M. (2020). Análisis de Sentimiento para Tweets en Español: Algunos Experimentos con Foco en los Tweets Neutros. *Procesamiento del Lenguaje Natural*, 64, 109-116.
- Daga, I., Gupta, A., Vardhan, R., & Mukherjee, P. (2020). Prediction of Likes and Retweets Using Text Information Retrieval. *Procedia Computer Science*, 168, 123-128.
- de Chernatony, L. (2001). *From Brand Vision to Brand Evaluation. Strategically Building and Sustaining Brands*. Oxford: Butterworth-Heinemann.
- Facebook (s.f.). <https://www.facebook.com/> [Consultado el 3/6/2021]
- Falk, R. F., & Miller, N. B. (1992). *A primer for soft modelling*. Ed. Univ of Akron Pr, Akron, Ohio.
- Fiesler, C., & Proferes, N. (2018). "Participant" perceptions of Twitter research ethics. *Social Media+ Society*, January-March, 1-14.
- Fishbein, M., & Ajzen, I. (1975), *Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research*. Addison-Wesley, Reading, MA
- Fombrun, C. J. (1996). Reputation: Realizing Value from the Corporate Image. *Harvard Business School Press*, Boston.
- Gálvez-Rodríguez, M. D. M., Caba-Pérez, C., & López-Godoy, M. (2016). Drivers of Twitter as a strategic communication tool for non-profit organizations. *Internet Research*, 26(5), pp. 1052-1071
- GitHub (2021). <https://github.com/jmpina/TwitterAccountsSentiment> [Consultado el 5/6/2021]
- Guerra-Velasco, L. P. (2008, 7 de mayo). Primeros pasos con Knime. Disponible en: http://laurel.datsi.fi.upm.es/_media/docencia/cursos/inap/ejemplodm.pdf [Consultado el 1/6/2021]
- Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, J. (2016). *Natural Language Processing: Python and NLTK*. Birmingham, UK: Packt Publishing.

Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media*, Ann Arbor, MI, June 2014.

IAB Spain (2020). *Estudio de Redes Sociales 2020*. Interactive Advertising Bureau. Disponible en:

<https://iabspain.es/download/39688/> [Consultado el 2/6/2021]

IBM SPSS. <https://www.ibm.com/analytics/spss-statistics-software> [Consultado el 1/6/2021]

Instagram (s.f.). <https://www.instagram.com/> [Consultado el 3/6/2021]

Jiménez Cano, R. (2017, 27 de septiembre). Twitter amplía el límite a 280 caracteres por mensaje. El País. Disponible en:

https://elpais.com/tecnologia/2017/09/26/actualidad/1506399545_718901.html

[Consultado el 1/6/2021]

Kapoor, P. S., Balaji, M. S., & Jiang, Y. (2021). Effectiveness of sustainability communication on social media: role of message appeal and message source. *International Journal of Contemporary Hospitality Management*, 0959-6119, 1-24.

KNIME (s.f.). <https://www.knime.com/> [Consultado el 1/6/2021]

Kuzma, J., Philippe, T. W., Koehler, J. W., & Coronel, R. S. (2016). Effects of Social Media Platform Preference: A Country Comparison of Online University Reputations. *Journal of International Management Studies*, 16(1), 35-44.

Kwak, H., Lee, C., Park, H. & Moon, S. (2010). What is Twitter, a Social Network or a News Media? *International World Wide Web Conference Committee*, 591-600.

Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and research: a systematic literature review through text mining. *IEEE Access*, 8, 67698-67717.

Li, M., Ch'ng, E., Chong, A. Y. L., & See, S. (2018). Multi-class Twitter sentiment classification with emojis. *Industrial Management & Data Systems*, 118 (9), 1804-1820.

Li, M., Chng, E., Chong, A. Y. L., & See, S. (2019). An empirical analysis of emoji usage on Twitter. *Industrial Management & Data Systems*, 119 (8), 1748-1763.

Liu, X., Burns, A. C., & Hou, Y. (2017). An investigation of brand-related user-generated content on Twitter. *Journal of Advertising*, 46(2), 236-247.

Llorente-Ayuso, P., Alarcón, D. I., & Royo, C. V. (2019). Análisis de sentimientos aplicado a la opinión política en Twitter: un sistema de clasificación en tiempo real. *Universitat Oberta de Catalunya*. Disponible en:

<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/96770/6/pllorenteayusoTFG0619memoria.pdf> [Consultado el 6/6/2021]

Liu, X. (2019). A big data approach to examining social bots on Twitter. *Journal of Services Marketing*, 33(4), 369-379.

Lovejoy, K., Waters, R. D., & Saxton, G. D. (2012). Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less. *Public Relations Review*, 38(2), 313-318.

McCallum, D.R., & Peterson, J.L. (1982). Computer-Based Readability Indexes. *Proceedings of the ACM '82 Conference*, October, 44-48.

- McCormick, T. H., Lee, H., Cesare, N., Shojaie, A., & Spiro, E. S. (2017). Using Twitter for demographic and social science research: Tools for data collection and processing. *Sociological Methods & Research*, 46(3), 390-421.
- McShane, L., Pancer, E., Poole, M., & Deng, Q. (2021). Emoji, playfulness, and brand engagement on twitter. *Journal of Interactive Marketing*, 53, 96-110.
- Morrissey, M., Wasser, L., & Farmer, C. (2020). Lesson 2. Automate Getting Twitter Data in Python Using Tweepy and API Access. Disponible en: <https://www.earthdatascience.org/courses/use-data-open-source-python/intro-to-apis/twitter-data-in-python/> [Consultado el 4/6/2021]
- MrExcel.com (2011, 22 de julio). VBA Count Syllables. Disponible en: <https://www.mrexcel.com/board/threads/vba-count-syllables.566173/> [Consultado el 1/6/2021]
- Noor, S., Guo, Y., Shah, S. H. H., Nawaz, M. S., & Butt, A. S. (2020). Research synthesis and thematic analysis of twitter through bibliometric analysis. *International Journal on Semantic Web and Information Systems*, 16(3), 88-109.
- Open-Office.es (2015, 24 de marzo). Contar las sílabas de una celda. Disponible en: <https://blog.open-office.es/calc/contar-las-silabas-de-una-celda> [Consultado el 1/6/2021]
- Pla, F., & Hurtado, L. F. (2018). Spanish sentiment analysis in Twitter at the TASS workshop. *Language Resources and Evaluation*, 52(2), 645-672.
- Prada, A., & Iglesias, C. A. (2020). Predicting reputation in the sharing economy with twitter social data, *Applied Sciences*, 10(8), 2881, 1-18.
- QQ (s.f.). <http://qq.com/> [Consultado el 3/6/2021]
- Qualman, E. (2012). *Socialnomics: How social media transforms the way we live and do business*. John Wiley & Sons.
- Roesslein, J. (2021). Tweepy Documentation. Release 3.10.0. Disponible en: https://docs.tweepy.org/_/downloads/en/stable/pdf/ [Consultado el 4/6/2021]
- Singhai, G. (2020, 1 de julio). Building a Twitter Sentiment Analysis in Python. Pluralsight. Disponible en: <https://www.pluralsight.com/guides/building-a-twitter-sentiment-analysis-in-python> [Consultado el 1/6/2021]
- Schlagwein, D., & Hu, M. (2017). How and why organisations use social media: five use types and their relation to absorptive capacity. *Journal of Information Technology*, 32(2), 194-209.
- Stewart, B., & Quan-Haase, A. (2017). Twitter as method: Using Twitter as a tool to conduct research. *The SAGE Handbook of Social Media Research Methods*, 251-265.
- Stone, J. A., & Can, S. H. (2020). Linguistic analysis of municipal twitter feeds: Factors influencing frequency and engagement. *Government Information Quarterly*, 37(4), 101468.
- Supeuser.com (2017, 15 de octubre). How to remove all emojis from Excel. Disponible en: <https://superuser.com/questions/1259236/how-to-remove-all-emojis-from-excel> [Consultado el 1/6/2021]

Swani, K., Brown, B. P., & Milne, G.R. (2014). Should tweets differ for B2B and B2C? An analysis of Fortune 500 companies' Twitter communications. *Industrial Marketing Management*, 43(5), 873-881.

Tabachnick, B.G. y Fidell, L.S. (1996). *Using Multivariate Statistics (4th ed.)*, Harper and Row, New York.

TASS (s.f.). TASS: Workshop on Semantic Analysis at SEPLN. <http://tass.sepln.org/> [Consultado el 1/6/2021]

Teh, P. L., Rayson, P., Pak, I., & Piao, S. (2015). Sentiment analysis tools should take account of the number of exclamation marks!!! *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, 1-6.

Tellis, G. J., MacInnis, D. J., Tirunillai, S., & Zhang, Y. (2019). What drives virality (sharing) of online digital content? The critical role of information, emotion, and brand prominence. *Journal of Marketing*, 83(4), 1-20.

TextBlob (s.f.). <https://textblob.readthedocs.io/> [Consultado el 1/6/2021]

TikTok (s.f.). <https://www.tiktok.com/> [Consultado el 3/6/2021]

Tsugawa, S., & Ohsaki, H. (2017). On the relation between message sentiment and its virality on social media. *Social Network Analysis and Mining*, 7(19), 1-19.

Tumblr (s.f.). <https://www.tumblr.com/> [Consultado el 3/6/2021]

Twitch (s.f.). <https://www.twitch.tv/> [Consultado el 3/6/2021]

Twitter (s.f.). <http://twitter.com/> [Consultado el 1/6/2021]

Twitter (s.f.). Twitter Developer. <https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview> [Consultado el 1/6/2021]

Universidad de Zaragoza (2010). *Manual de Identidad Corporativa*. Vicerrectorado de Relaciones Institucionales y Comunicación. Disponible en: <https://www.unizar.es/identidad-corporativa/identidad-corporativa> [Consultado el 1/6/2021]

Universidad de Zaragoza (2015). *Carta de Servicios*. Gabinete de Imagen y Comunicación. Disponible en: https://documenta.unizar.es/share/s/EDFzdDvkT1iJd9f9Xj_TsA [Consultado el 1/6/2021]

Universidad de Zaragoza (s.f.). Folleto Informativo. Gabinete de Rector. Disponible en: <http://unizar.es/institucion/conoce-la-universidad/folleto-informativo> [Consultado el 1/6/2021]

Universidad de Zaragoza (s.f.) Datos abiertos (Opendata). Disponible en: <https://zaguan.unizar.es/collection/opendata> [Consultado el 1/6/2021]

Universidad de Zaragoza (s.f.). <https://www.unizar.es/> [Consultado el 1/6/2021]

Universidad de San Jorge (s.f.). <https://www.usj.es/> [Consultado el 1/6/2021]

Urdaneta-Fernández, L.A. (2019, 4 de mayo). Reducir el número de palabras de un texto: lematización y radicalización (stemming) con Python. Medium. Disponible en: <https://medium.com/qu4nt/reducir-el-n%C3%BAmero-de-palabras-de-un-texto-lematizaci%C3%B3n-y-radicalizaci%C3%B3n-stemming-con-python-965bfd0c69fa> [Consultado el 1/6/2021]

- Vogler, D. (2020). Analyzing reputation of Swiss universities on Twitter—The role of stakeholders, content and sources. *Corporate Communications: An International Journal*, 25(3), 429-445.
- Von Hippel, P. T. (2004). Biases in SPSS 12.0 missing value analysis. *The American Statistician*, 58(2), 160-164.
- Wang, R., Liu, W., & Gao, S. (2016). Hashtags and information virality in networked social movement: Examining hashtag co-occurrence patterns. *Online Information Review*, 40(7), 850-866.
- Wang, Y., & Yang, Y. (2020). Dialogic communication on social media: How organizations use Twitter to build dialogic relationships with their publics. *Computers in Human Behavior*, 104, 106183.
- Wagner, T. F., Baccarella, C. V., & Voigt, K. I. (2017). Framing social media communication: Investigating the effects of brand post appeals on user interaction. *European Management Journal*, 35(5), 606-616.
- Wang, Y., & Zhou, S. (2015). How do sports organizations use social media to build relationships? A content analysis of NBA clubs' Twitter use. *International Journal of Sport Communication*, 8(2), 133-148.
- Wajid, A., Raziq, M. M., Ahmed, Q. M., & Ahmad, M. (2021). Observing viewers' self-reported and neurophysiological responses to message appeal in social media advertisements. *Journal of Retailing and Consumer Services*, 102373, 1-10.
- We Are Social & Hootsuite (2021). Digital 2021: The Latest Insights Into the 'State of Digital'. Disponible en: <https://datareportal.com/reports/digital-2021-global-overview-report> [Consultado el 1/6/2021]
- WeChat (s.f.). <https://www.wechat.com/> [Consultado el 3/6/2021]
- Westerman, D., Spence, P. R., & Van Der Heide, B. (2012). A social network as information: The effect of system generated reports of connectedness on credibility on Twitter. *Computers in Human Behavior*, 28(1), 199-206.
- Wilcox, D., Cameron, G., & Xifra, J. (2012). *Relaciones públicas. Estrategias y Tácticas (10ª Ed.)*. Pearson Addison Wesley, Madrid.
- Yumi (2018, 16 de enero). Extract someone's tweet using tweepy. Disponible en: <https://fairyonice.github.io/extract-someones-tweet-using-tweepy.html> [Consultado el 4/6/2021]
- Zhang, H., Sun, J., Liu, F., & Knight, J. G. (2014). Be rational or be emotional: advertising appeals, service types and consumer responses. *European Journal of Marketing*, 48 (11/12), 2105-2126.
- Zhang, H., Wang, P. J. J., & Chen, X. (2017). A novel decision support model for satisfactory restaurants utilizing social information: a case study of TripAdvisor.com. *Tourism Management*, 59, 281-297.
- Zimmer, M., & Proferes, N.J. (2014). A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66 (3), 250-261.

Anexo I. Diccionario de términos técnicos

- **Algoritmo de clasificación:** algoritmo de aprendizaje computacional utilizado para predecir la clase a la que pertenece un elemento del conjunto de datos.
- **Análisis de sentimientos:** disciplina de la inteligencia artificial encargada del estudio computacional de las opiniones, emociones o sentimientos expresados en el texto.
- **API** (*Application Programming Interface*): interfaz de programación de aplicaciones. Conjunto de clases, funciones o procedimientos que ofrece una biblioteca para ser usados por otro software.
- **Accuracy** (exactitud): métrica de evaluación de algoritmos de aprendizaje supervisado que se calcula como la proporción de predicciones correctas (verdaderos positivos y verdaderos negativos) sobre el total (suma de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos). Un valor próximo a 1 indica una mejor clasificación.
- **Aprendizaje automático** (*machine learning*): disciplina de la inteligencia artificial dedicada al desarrollo de algoritmos de aprendizaje para los computadores.
- **Conjunto de entrenamiento** (*training*): fase del proceso de aprendizaje supervisado en la que un algoritmo es ajustado según un conjunto de datos.
- **Conjunto de evaluación** (*evaluation*): fase del proceso de aprendizaje supervisado en la que se examina el poder predictivo del modelo previamente ajustado con el conjunto de entrenamiento, pero utilizando otros datos.
- **F-Score** (valor-F): métrica de evaluación de algoritmos de aprendizaje supervisado que combina *precision* y *recall* en una media armónica ponderada. Un valor próximo a 1 indica una mejor clasificación.
- **Falsos positivos:** cantidad de elementos clasificados incorrectamente como pertenecientes a una clase determinada.
- **Falsos negativos:** cantidad de elementos clasificados incorrectamente como no pertenecientes a una clase determinada.
- **Lenguaje natural:** lenguaje de propósito general que surge espontáneamente de la actividad comunicativa de los seres humanos.
- **Matriz de confusión:** matriz que representa los resultados de una clasificación. En las filas se muestran las clases reales a las que pertenecen los elementos de un conjunto de datos y, en sus columnas, las clases predichas para esos mismos elementos.
- **OAuth:** estándar abierto para la emisión segura de autorizaciones de comunicación con una API.
- **Python:** lenguaje de programación muy utilizado para proyectos de minería de datos y de *machine learning*.
- **Recall** (sensibilidad): métrica de evaluación de algoritmos de aprendizaje supervisado que se calcula como la proporción entre los elementos clasificados correctamente como pertenecientes a una clase (verdaderos positivos) y el total de elementos que pertenecen a dicha clase (verdaderos positivos y falsos negativos). Un valor próximo a 1 indica una mejor clasificación.
- **Stemming:** proceso por el cual una palabra es sustituida por su raíz eliminando sus afijos.
- **Stop Word:** palabra que no aporta significado a la oración.
- **Tokenización:** proceso de división del texto de un tuit en el conjunto de sus palabras.
- **Validación cruzada:** técnica de evaluación de los algoritmos de clasificación que divide el conjunto de datos de entrenamiento en K particiones de manera aleatoria y realiza una evaluación k veces, tomando en cada una de ellas un conjunto de prueba diferente para evitar la dependencia entre éste y el conjunto de entrenamiento.
- **Verdaderos positivos:** cantidad de elementos clasificados correctamente como pertenecientes a una clase determinada.
- **Verdaderos negativos:** cantidad de elementos clasificados correctamente como no pertenecientes a una clase determinada.

Nota: definiciones adaptadas del glosario elaborado por Llorente-Ayuso et al. (2019, pp. 41-45).

Anexo II. Compromisos de calidad e indicadores del Gabinete de Comunicación (1 de 2)

Servicio prestado	Grupos de interés	Compromiso	Indicador	Frecuencia de medición	Estándar de calidad
Atención a las peticiones de difusión de los miembros de la comunidad universitaria	<ul style="list-style-type: none"> ● Miembros de la comunidad universitaria ● Medios de comunicación ● Sociedad en general 	Atender las peticiones de miembros de la comunidad universitaria, realizadas en el marco de la Universidad de Zaragoza, para la difusión de proyectos o actividades en medios de comunicación, informándoles sobre su procedencia o no procedencia como hecho o actividad noticiable y, en caso afirmativo, difundirlo.	<ul style="list-style-type: none"> ● Número de noticias y Convocatorias enviadas a los medios de comunicación ● Número de informaciones publicadas en los medios de comunicación 	Anualmente	100%
	<ul style="list-style-type: none"> ● Miembros de la comunidad universitaria 	Contestar en un plazo máximo de tres días a las consultas o propuestas sobre convocatorias de ruedas de prensa, informando acerca de su procedencia o no procedencia y sobre la fecha de celebración de la misma.	● Número de consultas recibidas	Anualmente	100%
Comunicación interna	<ul style="list-style-type: none"> ● Miembros de la Comunidad universitaria 	Remitir a todos los miembros de la comunidad universitaria que lo deseen, el dossier de prensa diario de la Universidad, todos los días laborales, antes de las 11h.	● Hora de envío	Anualmente	Proporción de días en que se envía antes de las 11h
		Enviar el boletín iUnizar a todos los miembros de la comunidad universitaria, todos los días laborales.	● Número de informaciones recogidas	Diario en días laborales	100%
		Mantenimiento de la información de los mupis.	● Número de informaciones recogidas	Diario en días laborales	Seguimiento.
Comunicación externa	<ul style="list-style-type: none"> ● Medios de comunicación 	Atender todas las consultas de los medios de comunicación.	● Porcentaje de atención a las demandas de los medios de comunicación	Trimestral	100%
Comunicación Interna y externa	<ul style="list-style-type: none"> ● Unidades de la Universidad 	Apoyar el proceso de migración a la nueva web.	● Número de unidades que han migrado a la nueva web	Anual	Seguimiento.
	<ul style="list-style-type: none"> ● Miembros de la comunidad universitaria ● Medios de comunicación ● Sociedad 	Actualizar permanentemente los contenidos de la web de la Universidad.	● Número de consultas por páginas	Anual	Incremento del número de consultas en un 5% anual

Fuente: Universidad de Zaragoza (2015)

Anexo II. Compromisos de calidad e indicadores del Gabinete de Comunicación (2 de 2)

Servicio prestado	Grupos de interés	Compromiso	Indicador	Frecuencia de medición	Estándar de calidad
Imagen corporativa	<ul style="list-style-type: none"> • Miembros de la comunidad universitaria 	Atender todas las consultas sobre la imagen corporativa de la Universidad y resolverlas en el plazo máximo de una semana.	<ul style="list-style-type: none"> • Porcentaje de atención a las demandas relativas a la imagen corporativa de la Universidad 	Anual	100%
Asesoría sobre publicidad	<ul style="list-style-type: none"> • Unidades de la Universidad 	Atender las solicitudes de asesoramiento sobre inserción publicitaria que se hagan por las unidades.	<ul style="list-style-type: none"> • Número de folletos, carteles y otros soportes publicitarios coordinados por el Gabinete 	Anual	Seguimiento.
		Asesoramiento para las inserciones publicitarias institucionales.	<ul style="list-style-type: none"> • Número de folletos, carteles y otros soportes publicitarios coordinados por el Gabinete 	Anual	Seguimiento.
Redes sociales	<ul style="list-style-type: none"> • Miembros de la comunidad universitaria • Sociedad en general 	Gestión de redes sociales.	<ul style="list-style-type: none"> • Número de seguidores en redes sociales («me gusta», «favoritos», «retuits»...). 	Permanente	Incremento del número de seguidores en un 10% anual.

Fuente: Universidad de Zaragoza (2015)

Anexo III. Organismos de la UZ con cuentas en redes sociales (1 de 3)

Universidad de Zaragoza (General)	Empresas constituidas spinoff y startup UZ
Asambleas/delegaciones estudiantes	BEOnChip SL
Asamblea de estudiantes Teruel	Bitbrain Technologies SL
Asociación de Estudiantes de Fisioterapia	Blipol RH Analytics
Consejo de estudiantes	Centro Clínico OMT-E Fisioterapia SLP
Delegación alumnos Ciencias	Click and plan
Delegación de Est. Facultad de Veterinaria y CTA	CONET Mascotas
Delegación de Filosofía y Letras	EBERS Medical Technology
Delegación EINA	Ebronautas SL
Delegación Fac. Ciencias de la Salud	Energy On Site Plus SL
Delegación Fac. Derecho	Epic Power SL
Delegación Fac. Economía y Empresa (FECHEM)	Erasmus Play
Delegación Facultad Educación	Esciencia, eventos científicos
LGTBIQ Unizar	GEEZAR Soluciones SL
Asignaturas	Geoartec Technical Solutions SL
Algoritmia básica - Especialidad en Computación	Geospatiumlab SL
Asignatura de Comunicación e Información	H2i Analytics
Asignatura e-marketing	Herizont
Asignatura Plan de marketing	Historia de Aragón
Bioinformática- Especialidad en Computación	HP Lab History & Development SL
EducaFísicaTE	Hunteet Creativos SL
Fotografía, Fotoperiodismo y Técnica Imagen	Idiogram Technologies SL
Campus	Interebook Solutions SL
Campus de Huesca	Kampal Data Solutions SL
Campus de La Almunia	Libelium Comunicaciones Inalámbricas Distribuidas
Campus de Teruel	Nanoimmunotech SL
Universidad de la Experiencia	nB nanoScale Biomagnetics SL
Cátedras	Neki Creativos SL
Cat. Aves Nobles-Aldelís	Newfood Development SL
Cat. Banco Santander Unizar	NIVEL-4 SLP
Cat. Casa Matachín	NYXELL S. Coop.
Cat. CEMEX Unizar	RemOT Technologies SL
Cat. Colegio Juan de Lanuza de Innovación Educativa	Roamgym Wellness and Sport SL
Cat. Cooperación al Desarrollo (CATCODES)	Scanner Patrimonio e Industria
Cat. Participación e Innovación Democrática	SMT4V Researching SL (Smarthings4Vision)
Cat. Desarrollo Directivo Femenino	TRASLUZ
Cat. Emprender	Tu perro es bienvenido
Cat. FERSA	Up4Data
Cat. Magapor	Iniciativas de estudiantes
Cat. Oviaragón - Pastores	Alumnos ADE
Cat. SAMCA	Alumnos de Periodoncia e Implantología Oral
Cat. Despoblación y Creatividad	Asociación Estudiantil de Ciencia Animal (AECA)
Cat. Territorio, Sociedad y Visualización Geográfica	CineClub Cerbuna
Cat. Urbaser Tecnologías Innovadoras	EntreMedios
Cat. Zaragoza Vivienda	Instituto de Ingeniería Eléctrica y Electrónica Unizar
Gestión de cátedras	Voluntariado Unizar
Centros	Innovación docente
Centro Universitario de la Defensa (CUD)	ID Termodinámica e Ingeniería Términa (EINA)
Centro Universitario de Lenguas Modernas (CULM)	PID Aprendizaje - Servicio
Escuela de Doctorado	Red Entornos uLearning en Educación Superior
Escuela de Ingeniería y Arquitectura	Institutos y grupos de investigación
Escuela Politécnica Superior de Huesca	BIFI- IUI Biocomputación y Física Sist. Complejos
Escuela Universitaria de Turismo de Zaragoza	Biomedical Signal Interpretation Comp. Simulation-I3A
Escuela Universitaria Politécnica de La Almunia	Centro Astropartículas y Física Altas Energías (CAPA)
Escuela Universitaria Politécnica de Teruel	CIRCE - IUI Mixto
Facultad de Bellas Artes de Teruel	Computational Hydraulics Group
Facultad de Ciencias	GENUD

Fuente: Universidad de Zaragoza (datos del mapa de redes sociales consultados el 19/1/2021)

Anexo III. Organismos de la UZ con cuentas en redes sociales (2 de 3)

<p>Facultad de Ciencias de la Salud Fac. Ciencias de la Salud y del Deporte de Huesca Fac. Ciencias Humanas y de la Educación de Huesca Facultad de Ciencias Sociales y del Trabajo Facultad de Ciencias Sociales y Humanas de Teruel Facultad de Derecho Facultad de Economía y Empresa Facultad de Educación Facultad de Empresa y Gestión Pública de Huesca Facultad de Filosofía y Letras Facultad de Medicina Facultad de Veterinaria Instituto de Ciencias de la Educación (ICE)</p>	<p>GEOT GICIDUZ Group in Product and Project Management Grupo Aragosaurus G.I. Análisis Empresarial y Competitividad G.I. CEMA G.I. Educación Comunicativa y Literaria Sociedad Info G.I. Educación Física y Promoción Actividad Física G.I. ENFYRED G.I. Fisioterapia G.I. Personalidad, Emoción y Salud G.I. InterGEDI G.I. 'Método' Unizar G.I. Sociedad, Incertidumbre y Creatividad UZ</p>
<p>Colectivos universitarios</p>	<p>G.I. Tissue Microenvironment lab Grupo de Mecánica Aplicada y Bioingeniería-I3A Grupo de Sistemas de Información Avanzados (IAAA) Grupo de Tecnologías Fotónicas (GTF) Grupo Era - Emoción, Regulación y Ajuste I3A - IUI en Ingeniería de Aragón IA2 - IUI Mixto Agroalimentario de Aragón ICMA - IUI Mixto de Ciencia de Materiales de Aragón INA - IUI de Nanociencia de Aragón Instituto de Nanociencia y Materiales de Aragón International Axion Observatory (IAXO) IPH- Instituto de Patrimonio y Humanidades ISQCH - IUI Mixto Síntesis Química IUCA - IUI en Ciencias Ambientales de Aragón IUMA - IUI en Matemáticas y Aplicaciones Lab. Eco. Circular, Bioeconomía e Innov. Ciudadana Laboratorio de Economía Social Laboratorio de Paleogenómica Laboratorio de Sociología Jurídica Lingüística, Lengua y Cultura Francesas Liquid Crystals&Polymers CLIP Group_LAB Psylex ZLC - Zaragoza Logistics Center (Instituto adscrito)</p>
<p>AEGEE Zaragoza Agraluz APUDEPA Asoc. Diseño Industrial Unizar Asoc. Estudiantes de Odontología Asoc. Veterinaria equina de la UZ Asociación de debates de la Universidad de Zaragoza Asociación de mayores de la U. de la Experiencia AUFA Somerondón Coro Cantatutti Coro de Ingenieros Zaragoza Coro Enarmonía Coro Universitario del Campus de Huesca CUD Unizar (Centro Universitario de la Defensa) CUEP (Colectivo de Estudiantes Progresistas) DADEXpress EDU (Estudiantes en Defensa de la Universidad) Estudiantes en Movimiento Aragón FdE Aragón Morfeo Teatro Personal Investigador en Formación Plataforma por la dignidad en la investigación UZ Púlsar. Asociación de software libre Purna estudiantil ReFyLuz UDE Aragón (Unión de Estudiantes) Universitarios con la infancia</p>	<p>Servicios universitarios</p> <p>Área de Cultura Área de Proyección Social e Igualdad Asesorías (jurídica, orientación de estudios...) Biblioteca de la Universidad de Zaragoza Biblioteca del Campus Universitario de Teruel Biblioteca Fac. Empresa y Gestión Pública de Huesca Biblioteca Hypatia de Alejandría Biblioteca María Moliner Casa del Estudiante Centro de Documentación Europea Consulta de Endocrinología para animales de compañía Cursos de Español como Lengua Extranjera Cursos extraordinarios de verano Herramientas TIC para estudiantes Herramientas TIC Unizar Hospital Veterinario Universidad de Zaragoza Instituto Confucio Unizar Laboratorio de Microscopías Avanzadas</p>
<p>Colegios mayores y residencias de estudiantes</p>	
<p>CMU Cardenal Xavierre CMU Josefa Segovia CMU La Anunciata CMU Miraflores CMU Pablo Serrano (Teruel) CMU Pedro Cerbuna CMU Peñalba CMU Ramón Acín (Huesca) CMU Santa Isabel CMU Virgen del Carmen</p>	
<p>Departamentos</p>	
<p>Área de Didáctica de las Matemáticas D. Ciencias de la Tierra D. Dirección de Marketing e Investigación de Mercados</p>	

Fuente: Universidad de Zaragoza (datos del mapa de redes sociales consultados el 19/1/2021)

Anexo III. Organismos de la UZ con cuentas en redes sociales (3 de 3)

D. Dirección y Organización de Empresas D. Geografía y Ordenación del Territorio D. Historia del Arte D. Informática e Ingeniería de Sistemas D. Ingeniería Química y Ciencias del Medioambiente Unidad predepartamental de Arquitectura	Museo de Ciencias Naturales Observatorio de Igualdad Oficina de Atención a la Diversidad (OUAD) Oficina de Software Libre (OSLUZ) Oficina Verde Prensas Universitarias de Zaragoza (PUZ)
Deporte	Radio Unizar SEMETA Servicio de Actividades Deportivas Servicio de Informática y Comunicaciones Servicio de Publicaciones Servicio General de Apoyo a la Investigación - SAI Tienda-librería Paraninfo Universa Universidad de Verano de Teruel Unizar ADD MOOC Unizar Saludable
Asoc. Rugby Veterinaria 'Borregos' Balonmano Fecem Ciencias Voleibol Masculino Club Baloncesto Femenino Club Deportivo Universitario Rugby Deportes Cerbuna Deportes Ciencias Deportes Derecho Deportes Economía y Empresa Deportes Filosofía Deportes Medicina Derecho Rugby Equipo de Fútbol Femenino Facultad de Derecho Fútbol sala Oscense Unizar Kyudo Universidad de Zaragoza Rugby Filosofía y Letras	Titulaciones Experto Universitario Gestión Responsabilidad Social Grado ADE Campus de Teruel Máster en Educación Socioemocional M. Gestión Administrativa M. Gestión de la Innovación ITA M. Gestión Internacional y Comercio Exterior M. Medicina de Urgencia y Rescate en Montaña M. Museos M. Musicoterapia M. Robotics, Graphics and Computer Vision M. MeBA Unizar M. Millennium en Derecho Internacional Privado M. Desarrollo Empresarial M. Dirección y Gestión Recursos y Servicios Sociales M. Lectura, libros y lectores infantiles y juveniles M. Economía Mención de Educación Física Fac. CCSSHH Teruel
Egresados	Vicedecanatos Actividades Culturales - FCSHT Vicedecanato de Estudiantes y Movilidad FCST Vicerrectorados Vicerrectorado de Estudiantes y Empleo Vicerrectorado Tecnologías INFOCOM Vicerrectorado del Campus de Huesca Vicerrectorado del Campus de Teruel
Servicios centrales	Otros Diseño de producto y marca Equipo docente de electromagnetismo Unizar FEUZ (Fundación Empresa Universidad) Geografialia Relaciones Internacionales EINA Relaciones internacionales FECOM Tropelias
Administración Electrónica Área de Calidad y Mejora Gabinete de Imagen y Comunicación OTRI Sección de Archivo Sección de Becas Secretaría General Unidad de Cultura Científica Unidad de Seguridad Unidad Técnica de Construcciones y Mantenimiento	
Secciones sindicales	
CGT La universidad que queremos Somos Unizar UGT Universidad de Zaragoza	

Fuente: Universidad de Zaragoza (datos del mapa de redes sociales consultados el 19/1/2021)

Anexo IV. Código para extraer tuits de varias cuentas simultáneamente

```
import tweepy
import csv
consumer_key='...'
consumer_secret='...'
access_token='...'
access_token_secret='...'
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)
f = open('listacuentas.txt', 'r')
users = f.read().splitlines()
f.close()
for user_name in users:
    all_tweets=[]
    tweets=api.user_timeline(screen_name = user_name, count=200, include_rts = False, tweet_mode =
'extended')
    all_tweets.extend(tweets)
    oldest = all_tweets[-1].id - 1
    while len(tweets) > 0:
        print ("getting tweets before %s" % (oldest))
        tweets = api.user_timeline(screen_name = user_name, count=200, include_rts = False, max_id=oldest,
tweet_mode = 'extended')
        all_tweets.extend(tweets)
        oldest = all_tweets[-1].id - 1
    from pandas import DataFrame
    outtweets = [[tweet.id_str, tweet.user.name, tweet.user.screen_name, tweet.user.description,
tweet.user.location, tweet.user.followers_count,
tweet.user.friends_count,tweet.user.favourites_count,tweet.user.statuses_count,tweet.user.listed_count,
tweet.user.default_profile,tweet.user.default_profile_image,tweet.created_at, tweet.favorite_count,
tweet.retweet_count, tweet.full_text.encode("utf-8").decode("utf-8")]
    for idx,tweet in enumerate(all_tweets)]
    df
=DataFrame(outtweets,columns=["id","name","sreen_name","description","location","followers_count","
friends_count","favourites_count","statuses_count",
"listed_count","default_profile","default_profile_image","created_at","favorite_count","retweet_count",
"text"])
    nombre="tuits" + user_name
    filename= "%s.csv" % nombre
    df.to_csv(filename,index=False)
    df.head(3)
```

Anexo V. Código para construir y evaluar el modelo de clasificación

```
import pandas as pd
import numpy as np
import re
import string
from nltk.tokenize import word_tokenize
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn.metrics import recall_score
from sklearn.metrics import precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import confusion_matrix
def load_dataset(filename, cols):
    dataset = pd.read_csv(filename, encoding = 'latin-1')
    dataset.columns = cols
    return dataset
def remove_unwanted_cols(dataset, cols):
    for col in cols:
        del dataset[col]
    return dataset
def tokenizer(tweet):
    tweet_tokens=word_tokenize(tweet)
    tokenized_words=[w for w in tweet_tokens]
    return " ".join(tokenized_words)
def get_feature_vector(train_fit):
    vector = TfidfVectorizer(sublinear_tf=True)
    vector.fit(train_fit)
    return vector
dataset = load_dataset("data/clas_modelo.csv", ['ID', 'SOCIAL', 'TEXTO'])
n_dataset = remove_unwanted_cols(dataset, ['ID'])
dataset.text = dataset['TEXTO'].apply(tokenizer)
dataset.text = dataset['TEXTO'].apply(tokenizer)
tf_vector = get_feature_vector(np.array(dataset.iloc[:, 1]).ravel())
X = tf_vector.transform(np.array(dataset.iloc[:, 1]).ravel())
y = np.array(dataset.iloc[:, 0]).ravel()
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
NB_model = MultinomialNB()
NB_model.fit(X_train, y_train)
y_predict_nb = NB_model.predict(X_test)
print(accuracy_score(y_test, y_predict_nb))
print(precision_score(y_test, y_predict_nb))
print(recall_score(y_test, y_predict_nb))
print(f1_score(y_test, y_predict_nb))
print(confusion_matrix(y_test, y_predict_nb))
LR_model = LogisticRegression(solver='lbfgs')
LR_model.fit(X_train, y_train)
y_predict_lr = LR_model.predict(X_test)
print(accuracy_score(y_test, y_predict_lr))
print(precision_score(y_test, y_predict_lr))
print(recall_score(y_test, y_predict_lr))
print(f1_score(y_test, y_predict_lr))
print(confusion_matrix(y_test, y_predict_lr))
```