Clasificación automática de textos sobre Trastornos de Conducta Alimentaria (TCA) obtenidos de Twitter

JOSÉ ALBERTO BENÍTEZ ANDRADES

MÁSTER EN INGENIERÍA Y CIENCIA DE DATOS
UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

Trabajo Fin Máster en Ingeniería y Ciencia de Datos

25 de junio de 2021

Directores:

Rafael Pastor Vargas
María Teresa García Ordás

Resumen

Gran parte de nuestra sociedad da mucha importancia al aspecto físico. Tener un aspecto físico bonito, o ideal, no tiene por qué ser sinónimo de padecer una buena salud. Hay personas que pueden poseer un cuerpo delgado, fuerte y con una buena definición muscular, pero no mantener una vida muy saludable debido al uso de sustancias que no son precisamente sanas. En la búsqueda de un aspecto físico delgado, muchos piensan que también encontrarán la felicidad y, sin embargo, en muchas ocasiones esto no es así. La presión que imponen los cánones de belleza extendidos en la sociedad han provocado que gran parte de la sociedad padezca enfermedades mentales relacionados con la alimentación y más conocidos como trastornos de conducta alimentaria (TCA). Algunos ejemplos son la anorexia, la bulimia, la bulimarexia, trastornos por atracón, etc.

Los medios sociales y, más concretamente, las redes sociales, son herramientas tecnológicas que cada vez tienen un mayor número de usuarios y mediante los cuales es sencillo difundir y divulgar información entre la sociedad de forma rápida. A pesar de las bondades de las redes sociales, cuando una persona padece un TCA, pueden fomentar o provocar un empeoramiento en la propia enfermedad del paciente. Twitter es una de las redes sociales más utilizadas en el ámbito de la recopilación de datos para su posterior estudio y, en el ámbito de la salud, es la herramienta más utilizada en el ámbito de la investigación en lo que a redes sociales se refiere. Son muchos los estudios que están centrándose en obtener tuits y generar modelos capaces de clasificar textos o analizar el sentimiento en distintas temática sanitarias.

Por todo lo anteriormente expuesto, en esta investigación se ha propuesto aplicar distintas técnicas de minería de datos recopilando distintos tuits relacionados con TCA. Posteriormente, se han aplicado técnicas de minería de textos y procesamiento de lenguaje natural que han permitido generar modelos predictivos haciendo uso de distintas técnicas de aprendizaje automático supervisado como bosques aleatorios, redes neuronales recurrentes e incluso modelos conocidos como Bidirectional Long Short-Term Memory (Bi-LSTM). Tras la aplicación de estos modelos, se ha puesto en valor la exactitud obtenida de los diferentes modelos de clasificación para clasificar tuits relacionados con TCA dentro de cuatro categorías diferentes que pueden ser de interés para la comunidad científica: (i) mensajes escritos por personas que padecen, o no, TCA, (ii) mensajes que fomentan, o no, el padecer un TCA, (iii) mensajes de carácter informativo, o de opinión y (iv) mensajes de carácter científico, o no.

Palabras clave

Minería de datos, minería de textos, procesamiento de lenguaje natural, redes neuronales, BERT.

Abstract

Much of our society attaches great importance to physical appearance. Having a beautiful, or ideal, physical appearance is not necessarily synonymous with good health. There are people who may have a slim, strong body with good muscle definition, but do not maintain a very healthy lifestyle due to the use of substances that are not exactly healthy. In the pursuit of a slim physique, many think that they will also find happiness, and yet this is often not the case. The pressure imposed by the canons of beauty prevalent in society has led to a large part of society suffering from mental illnesses related to eating, better known as eating disorders (ED). Some examples are anorexia, bulimia, bulimarexia, binge eating disorders, etc.

Social media and, more specifically, social networks, are technological tools that have an increasing number of users and through which it is easy to disseminate and spread information among society quickly. Despite the benefits of social networks, when a person suffers from an ED, they can promote or provoke a worsening of the patient's own illness. Twitter is one of the most widely used social networks in the field of data collection for subsequent study and, in the field of health, it is the most widely used tool in the field of research as far as social networks are concerned. Many studies are focusing on obtaining tweets and generating models capable of classifying texts or analysing sentiment in different health topics.

For all of the above reasons, this research has proposed to apply different data mining techniques by collecting different tweets related to ATT. Subsequently, text mining and natural language processing techniques have been applied to generate predictive models using different supervised machine learning techniques such as random forests, recurrent neural networks and even models known as Bidirectional Long Short-Term Memory (Bi-LSTM). After the application of these models, the accuracy obtained from the different classification models has been assessed to classify ATT-related tweets into four different categories that may be of interest to the scientific community: (i) messages written by people with or without ED, (ii) messages that do or do not encourage ED, (iii) messages of an informative or opinionated nature, and (iv) messages of a scientific or non-scientific nature.

Keywords

Data mining, text mining, natural language processing, neural networks, BERT.

Índice general

Índice de Figuras

Índice de Tablas

Agradecimientos

Este trabajo de fin de máster se presenta como la culminación de dos años de esfuerzo que no han sido para nada fáciles. Compatibilizar mi día a día en la Universidad, como personal docente e investigador, con todas las tareas que ello conlleva, junto con el resto de actividades que generalmente realizo, no ha sido fácil. Aún así, este documento refleja el fin de este sobreesfuerzo de dos años en el que he tenido que hipotecar muchos fines de semana y días festivos para poder finalizarlo.

No quiero finalizar este trabajo sin dar las gracias a todos aquellos que día a día me ayudan a percibir la vida con alegría y a seguir remando sin desistir en conseguir todo aquello que busco sin desfallecer en el intento y sin perder la sonrisa. Por todo ello, doy las gracias a mi novia, Camino, porque si no fuera como es, hoy en día no habría conseguido muchas de las cosas que realmente quería conseguir. Gracias por permitirme ser yo mismo. Agradezco a mis padres, a mi hermano, mi cuñada y mis sobrinos su comprensión constante, porque al final todas estas metas que me pongo, me restan tiempo que podría pasar también con ellos. No podría faltar mi agradecimiento a Lyu, mi gato, el cual me da un cariño incondicional desde hace unos años y que nunca me falla. Agradezco a mis socios Jesús y Cris su apoyo continuo siempre. Y agradecimiento enorme a todos mis compañeros de la Universidad de León con los que día a día conseguimos que trabajar sea más que eso y que, más que compañeros, realmente somos prácticamente amigos, porque somos una piña y nos apoyamos en todo: Isaías, Carmen, Héctor, Alija, Maite, Ángel. También le agradezco enormemente todo el apoyo a mis compañeras y compañeros del grupo de investigación SALBIS y, con el permiso del resto, especialmente a Pilar.

Por último, agradecer a todo el personal docente de este Máster lo que nos han enseñado en las distintas asignaturas, gracias a las cuales ha sido posible realizar este trabajo final. Con permiso del resto de docentes, mi especial agradecimiento a Rafael por dirigirme este trabajo junto a mi compañera y amiga Maite, que ya la mencioné anteriormente.

Capítulo 1

Introducción

Este capítulo se divide principalmente en tres subsecciones: antecedentes y contexto, objetivos generales y específicos y estructura de la memoria. En cada uno de ellos se detallará toda la información relacionada con la introducción de este trabajo de investigación.

1.1. Antecedentes y contexto

Una gran parte de la sociedad actual da una gran importancia al aspecto físico. Este hecho provoca que muchas personas dediquen bastante tiempo y esfuerzo en mejorar su apariencia, especialmente las mujeres (Harris & Carr, 2001; Senín-Calderón et al. 2020). Algunos de los mensajes enviados por revistas y medios de comunicación suelen incluir ideas que inducen a comprar productos relacionados con la estética y la belleza entre los que se incluyen ropa (siguiendo ciertas modas), cremas antiarrugas, y, sobre todo, productos que ayuden a adelgazar o perder peso indicando que una persona delgada será más feliz (Lou & Tse, 2020). Un ejemplo del arraigo tan fuerte que existe a este binomio de delgadez y felicidad son los debates continuos que han surgido y surgen en la actualidad acerca de las personas que trabajan como modelo para distintas campañas publicitarias. En contraposición, ha surgido un grupo de población que parece fomentar las modelos *curvy* (Izquierdo et al. 2019; Urdapilleta et al. 2019).

El exceso de peso durante la infancia y la adolescencia sigue siendo uno de los problemas más importantes de la salud mundial, a pesar de que surgió como una preocupación hace

varias décadas (Lobstein et al. 2004; Lobstein et al. 2015). Algunas estimaciones actuales sugieren que, aproximadamente, 40 millones de niños menores de 5 años y más de 330 millones de niños y adolescentes de 5 a 19 años tenían sobrepeso o eran obesos en 2016 ("Home - Global Nutrition Report", s.f.). Según la Organización Mundial de la Salud (OMS) en 2016, más de 1.900 millones de adultos de 18 años tenían sobrepeso, lo que representa un 39 % de la población adulta mundial. De ellos, más de 650 millones de adultos, un 15 % de la población eran obesos. La prevalencia mundial de la obesidad casi se triplicó entre 1975 y 2016 (World Health Organization, 2020).

Atendiendo a esta información, es posible destacar que gran parte de la población se beneficiaría de comer mejor y de realizar una mayor actividad física para mejorar la salud y la forma física. El simple hecho de observar lo que comemos y ser conscientes de los ingredientes de un producto o su composición nutricional, no es un trastorno de conducta alimentaria (TCA). Los TCA son enfermedades potencialmente mortales que son simultáneamente de naturaleza psicológica y física (Griffen et al. 2018). Se caracterizan por una serie de conductas alimentarias anormales y perjudiciales que van acompañadas y motivadas por creencias, percepciones y expectativas poco saludables en relación con la alimentación, el peso y la forma corporal (Griffen et al. 2018; Hoek, 2016). Como caracterización general, los individuos con TCA tienden a tener dificultades para aceptar y sentirse bien consigo mismos. Tienden a pensar que son "gordos" y "feos" debido al tamaño y la forma de su cuerpo, incluso cuando esta autocrítica es objetivamente inexacta y falsa. Al identificarse y definirse según su "gordura" percibida, las personas con trastornos alimentarios tienden a concluir que son inaceptables e indeseables y, como resultado, se sienten bastante inseguras e inadecuadas, especialmente en relación con sus cuerpos. Para ellos, controlar sus conductas alimenticias es el camino lógico en su búsqueda de la delgadez (Griffen et al. 2018; Hoek, 2016).

Además, a pesar de la complejidad de integrar todos los datos de prevalencia de los TCA, los estudios más recientes confirman que los TCA tienen una alta prevalencia en todo el mundo, especialmente en las mujeres. Además, la media ponderada de la prevalencia

puntual de TCA aumentó durante el período de estudio del 3,5 % para el período 2000-2006 al 7,8 % para el período 2013-2018. Esto pone de manifiesto un verdadero reto para la salud pública y los proveedores de servicios de salud (Galmiche et al. 2019).

Desde hace años, con la aparición de las redes sociales digitales como Twitter, Facebook e Instagram, entre otras, los medios de comunicación social se utilizan también como un instrumento de investigación para analizar debates públicos sobre temas muy diversos (Ackland, 2009; Carceller-Maicas, 2016; Lopez-Castroman et al. 2020; Timmins et al. 2018). Gracias a ellos y a la aparición de las estrategias de gestión y procesamiento de datos mediante técnicas de inteligencia artificial (IA) como, por ejemplo, la minería de textos, el aprendizaje automático o las técnicas de *deep learning*, estos medios pueden ayudar a analizar cuestiones de atención de la salud, como la obesidad, los trastornos de conducta alimentaria y derivados, desde la perspectiva del paciente (Bauer & Lizotte, 2021; Fiumara. et al. 2018; Musacchio et al. 2020). Estas investigaciones son posibles gracias a que las redes sociales promueven una cultura participativa que fomenta la creación de comunidades en las que se comparte información. Esto permite que puedan utilizarse de forma eficaz con fines científicos. Además, estos medios sociales cada vez son más populares, aumentando así su utilidad como fuentes para la investigación en materias de salud (Arigo et al. 2018).

Según diferentes estudios, es posible asegurar que los medios de comunicación social son, actualmente, una de las fuentes de datos que pueden ayudar a la investigación en el ámbito de la salud. Las investigaciones en salud que hacen uso de datos procedentes de medios sociales están en continuo crecimiento y se dividen principalmente en dos dominios: recopilación y predicción en tiempo real de enfermedades y recopilación de datos generados por usuarios y pacientes en los medios sociales. Gracias a los medios sociales es posible realizar estudios de observación que pueden ser utilizados para investigar acerca de la información que comparten los usuarios sobre un tema particular. Entre las principales plataformas de medios sociales existentes y que son utilizadas para la investigación de la salud, las más destacadas son: Twitter, Instagram, Youtube, Facebook, Reddit y otro blogs y foros intere-

santes (Timmins et al. 2018). Debido a su versatilidad y a la facilidad de recopilación y procesamiento de los datos que se obtienen, Twitter es la herramienta más utilizada para investigaciones relacionadas con la salud (Timmins et al. 2018). Los temas relacionados con la salud más investigados son aquellos relacionados con la organización de la asistencia sanitaria, la medicina del comportamiento, la psiquiatría, la neurología, las enfermedades infecciosas y la oncología (Sinnenberg et al. 2017).

Profundizando en los estudios que se han realizado sobre datos obtenidos de los medios sociales y los TCA, es posible encontrar investigaciones relacionadas con la detección de personas a favor y en contra de algunos TCA (Fettach & Benhiba, 2019; Lewis & Arbuthnott, 2012; Oksanen et al. 2015), detección de tuits informativos y no informativos (Viguria et al. 2020; Zhou et al. 2020), e incluso la detección de comunidades de personas con TCA (Fettach & Benhiba, 2019; Wang et al. 2018).

Sin embargo, no se han encontrado estudios que hagan uso del conjunto de datos para crear clasificadores que detecten (i) tuits que han sido escritos por personas que padecen o han padecido TCA, (ii) tuits que fomentan el padecer un TCA, (iii) tuits informativos o no informativos y, dentro de los tuits informativos, (iv) cuáles son aquellos que hacen uso de información de carácter científico y cuáles no.

1.2. Objetivos generales y específicos

Esta investigación surge de las siguientes cuestiones de investigación:

- **Pregunta 1:** ¿Es posible conseguir modelos de aprendizaje automático o aprendizaje profundo capaces de clasificar de forma precisa tuits sobre TCA en las cuatro categorías mencionadas en este trabajo?
- **Pregunta 2:** ¿Qué modelos de aprendizaje automático o aprendizaje profundo obtienen unos mejores resultados ante la problemática propuesta en esta investigación?

Los objetivos generales de este trabajo atienden a todas las competencias básicas, gene-

rales, transversales y específicas del Máster Universitario en Ingeniería y Ciencia de Datos. Estos objetivos son los expuestos a continuación:

- **Objetivo 1:** Aplicar técnicas de minería de datos que permitan generar un conjunto de datos para su posterior etiquetado y preprocesamiento.
- **Objetivo 2:** Generar modelos de aprendizaje automático y aprendizaje profundo capaces de clasificar textos sobre TCA con un alto grado de exactitud.
- **Objetivo 3:** Comparar entre los diferentes modelos cuál es el que mejores resultados ofrece en base a las categorías que se pretenden clasificar o predecir.

Atendiendo a una descripción más detallada de estos objetivos generales es posible destacar los siguientes objetivos específicos:

- **Objetivo específico 1:** Recopilar información de la red social Twitter aplicando distintas técnicas de minería de datos como, por ejemplo, *web scraping*.
- **Objetivo específico 2:** Aplicar técnicas de preprocesamiento de datos sobre los tuits recopilados que permitan su uso con diferentes técnicas de inteligencia artificial.
- **Objetivo específico 3:** Conseguir crear un subconjunto de tuits etiquetados en diferentes categorías que luego permitan entrenar distintos modelos de aprendizaje automático y aprendizaje profundo. Este subconjunto dispondrá de cuatro columnas binarias (sí/no) en base a si los tuits (i) han sido escritos por alguien que padece o ha padecido un TCA, (ii) promueven el padecer TCA, (iii) son de carácter informativo o (iv) son de carácter científico.
- **Objetivo específico 4:** Aplicar distintas técnicas de procesamiento de lenguaje natural (PLN) que den lugar a modelos predictivos eficientes dentro de las categorías clasificadas previamente.

- **Objetivo específico 5:** Comparar y mejorar los resultados de los modelos predictivos en base a los resultados obtenidos.

Los objetivos descritos convierten este trabajo de fin de máster en un trabajo de investigación integral de ingeniería aplicada al campo de la ciencia de datos en el que se sintetizan las competencias adquiridas en las enseñanzas. Además, involucran la coordinación entre los conocimientos, habilidades y destrezas que se han adquirido a lo largo de la formación en este Máster.

1.3. Estructura de esta memoria

La memoria se estructura de la siguiente forma:

- **Capítulo 1: Introducción**: Se trata del presente capítulo en el que se explican los antecedentes y el contexto del trabajo de investigación realizado, así como los objetivos esperados y la estructura de la memoria.
- **Capítulo 2: Estado del arte**: Se profundiza más en los diferentes estudios relacionados con los medios sociales, la informática sanitaria y los métodos de clasificación aplicados a problemas de salud.
- **Capítulo 3: Material y métodos**: Se expone la estructura del trabajo desarrollado exponiendo un resumen del flujo de trabajo realizado, la recopilación de tuits, el preprocesamiento de los datos y los modelos de clasificación aplicados al problema de estudio.
- **Capítulo 4: Experimentos y resultados**: Se detallan los aspectos metodólogicos aplicados al problema de estudio indicando los resultados obtenidos en la recopilación y preprocesamiento de datos, la implementación de los modelos de clasificación así como los resultados obtenidos con cada modelo en el conjunto de datos de estudio.

- **Capítulo 5: Discusión y conclusiones**: Se comentan los resultados obtenidos indicando las respuestas a las preguntas de investigación que fueron formuladas en el capítulo de introducción, así como la implicación de estos resultados, las limitaciones de investigación y las posibles líneas de investigación futuras.

- **Capítulo 6: Planificación y estimación de costes**: Se describe la planificación temporal de este trabajo junto con una estimación de los costes monetarios del mismo..

- **Capítulo 7: Aportaciones científicas**: Se detallan las aportaciones científicas que han surgido de esta investigación en forma de comunicación de congreso.

Capítulo 2

Estado del arte

Siguiendo los objetivos definidos en el capítulo 1, este capítulo expone los estudios más relevantes para la investigación propuesta. En primer lugar, se definen las redes sociales y se describen sus principales plataformas. A continuación, se analiza cómo se pueden utilizar los medios sociales con fines informáticos en el ámbito de la salud. Posteriormente, se describe el aprendizaje automático y se presentan los métodos de clasificación que se utilizan habitualmente en la investigación sanitaria. Tras este apartado, se resume cómo las redes sociales pueden ayudar en las investigaciones relacionadas con las TCA. Y, finalmente, se ofrece una síntesis de la literatura y se expone de forma razonada una justificación de la investigación realizada en este trabajo.

2.1. Medios sociales: definición y principales plataformas

La definición de medios de comunicación social o medios sociales no es trivial y, además, se convierten en una tarea compleja debido, entre otros motivos, a la rápida expansión tecnológica y a las diversas formas de comunicación existentes. Algunos investigadores coinciden en que los medios sociales comparten una serie de elementos comunes (Boyd & Ellison, 2007; Obar & Wildman, 2015; Sohota, 2020). En primer lugar, los medios sociales presentan unas aplicaciones basadas en la Web 2.0 de Internet que fomentan la participación de los usuarios. En segundo lugar, el contenido generado por el usuario es un componente esencial de todos los medios sociales. En tercer lugar, los individuos deben crear perfiles específicos

que son mantenidos por las plataformas de medios sociales. Por último, los usuarios pueden conectarse con otros individuos o con sus grupos en función de sus preferencias y de la información de su perfil. Los medios sociales presentan, pues, una tecnología interactiva en línea con contenidos generados por los usuarios. La característica más importante es que los medios sociales pueden estimular una cultura participativa a gran escala y pueden diferir en funcionalidad y alcance.

Las características descritas anteriormente han convertido a los medios sociales en un fenómeno tanto social como empresarial (Obar & Wildman, 2015) con más de dos mil quinientos millones de usuarios en todo el mundo (Statista, 2021). La posibilidad de crear una red en torno a intereses específicos ha facilitado el desarrollo de diferentes comunidades y ha dado lugar a la aparición de diversas plataformas de medios sociales (Vaganov et al. 2020). La figura 2.1 ofrece una visión general de las plataformas de medios sociales más populares clasificadas por el número de usuarios en millones. Atendiendo a la literatura científica y a las estadísticas actuales, es posible decir que que las plataformas de medios sociales actuales incluyen comunidades de interés, como las que se basan en un tipo específico de contenido, y aplicaciones de comunicación entre plataformas, como las aplicaciones de mensajería WhatsApp, Facebook y WeChat.

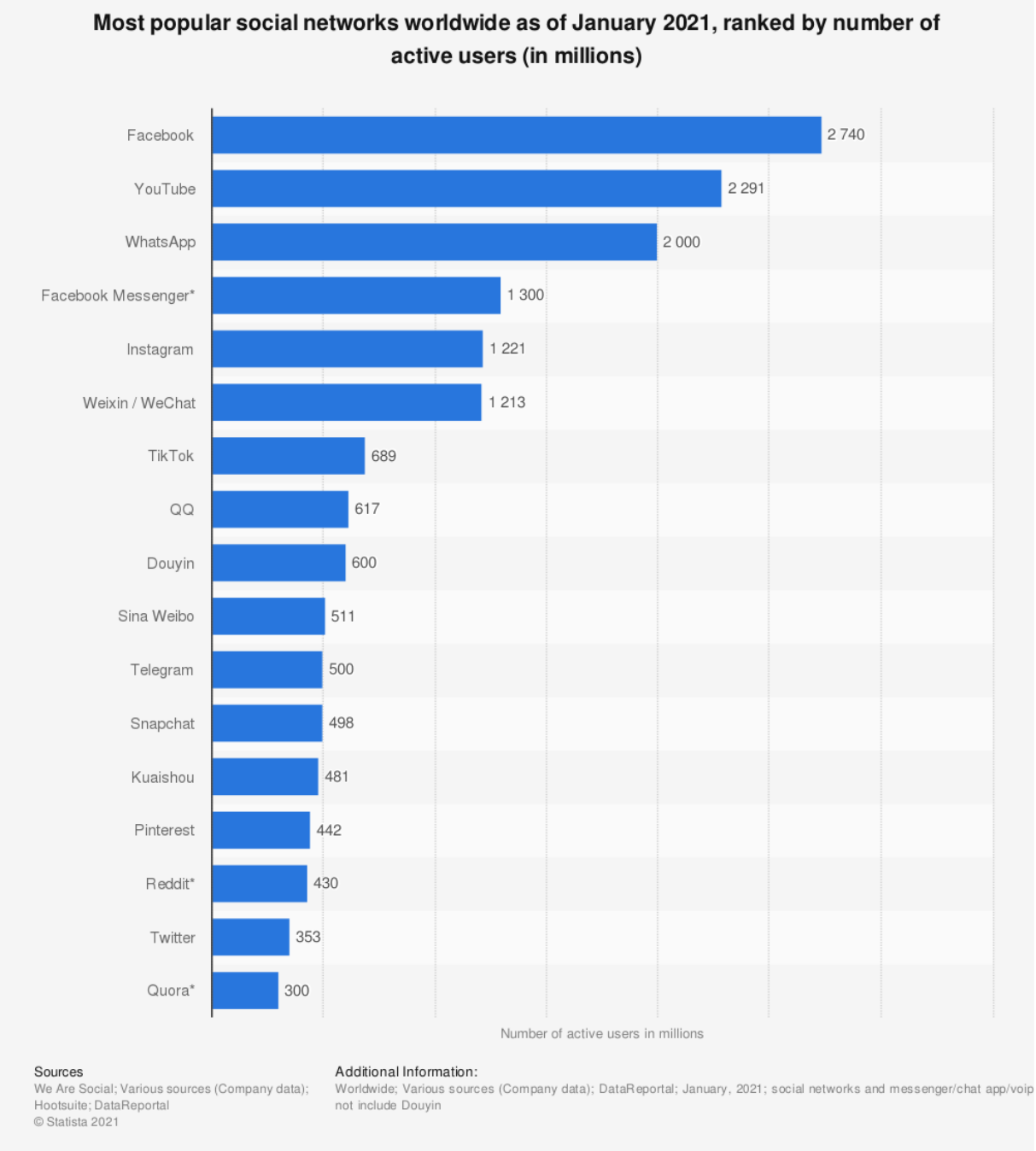

Figura 2.1: *Plataformas de medios sociales más populares en todo el mundo en enero de 2021, clasificadas por número de usuarios activos (en millones) en (Statista, 2021)*

En enero de 2021, la red social más popular es Facebook con 2.740 millones de usuarios en todo el mundo y le sigue YouTube con 2.291 millones (Statista, 2021). Twitter tiene solo 353 millones de usuarios (Statista, 2021). Sin embargo, junto con Facebook, estas plataformas son reconocidas como los medios sociales más destacados que emplean características de la Web 2.0 o de participación social en sus servicios (Martínez-Rojas et al. 2018). De este modo, fomentan la interacción entre los usuarios a través de los contenidos generados por ellos mismos. En consecuencia, con un número considerable de usuarios y contenidos compartidos, los medios sociales presentan fuente de información muy valiosa para la investigación.

Sin embargo, los estudios sobre los medios sociales deben tratarse con precaución, especialmente en lo que respecta a la generalización de los resultados. Hay que tener en cuenta que, al estar construidas en torno a intereses específicos, las plataformas de medios sociales tienden a atraer a usuarios con diferentes características sociales y demográficas (Zarrinkalam et al. 2020). La consideración más importante es que los usuarios de las redes sociales de una determinada plataforma no representan necesariamente a toda la población. Por ejemplo, una investigación de (Blank, 2017) determinó que los usuarios de Twitter procedentes de Reino Unido suelen caracterizarse por ser "más jóvenes, más ricos y más educados" en comparación con otros usuarios de Internet de Reino Unido. Coincide también que los usuarios de Twitter estadounidenses también son más jóvenes y más ricos, aunque no mejor educados. No obstante, estas apreciaciones o resultados encontrados en dicha investigación, no reflejan la sociedad completa en cada uno de estos países. Muy posiblemente haya un sesgo provocado por el hecho de que es más común que las personas más adineradas hagan uso de ciertos medios sociales.

Gracias al crecimiento de los medios sociales y, más concretamente, de las redes sociales, han aparecido nuevas oportunidades para recopilar datos para diferentes estudios, entre ellos, estudios relacionados con la salud (Mehmet et al. 2020; Mheidly & Fares, 2020; Stellefson et al. 2020). Cada vez son más las personas que utilizan los medios sociales para compartir información. Además, los medios sociales tienden a expandirse y a aumentar su impacto

global (Martín-Rojas et al. 2021). Por lo tanto, es importante entender cómo pueden, los medios sociales, formar y reflejar el discurso público.

La investigación propuesta en este trabajo hace uso de datos de medios sociales, en este caso, de la red social Twitter. Esta plataforma nacio en el año 2006 como una plataforma de microblogging y, a día de hoy, Twitter cuenta con 353 millones de usuarios en todo el mundo (Statista, 2021). En el año 2017, Twitter duplicó el límite de caracteres, pasando de los 140 iniciales a los 280 caracteres actuales. este cambio propulsó el uso de esta plataforma con fines científicos. Twitter es capaz de captar un debate público muy amplio y, gracias a ello, se ha convertido en una de las herramientas de investigación más utilizadas en las ciencias sociales y de la salud (Edo-Osagie et al. 2020; Jordan et al. 2019; Sinnenberg et al. 2017). Otra de las ventajas de Twitter es su fácil acceso a los datos en tiempo real mediante una interfaz de programación de aplicaciones (API). Gracias a la inmediatez y a la gran cantidad de datos, los expertos han podido recopilar con facilidad datos sobre diversos temas de interés para sus investigaciones. Además, la mayor parte de estudios de los medios sociales sobre TCA se basa en datos de Twitter (Chancellor & De Choudhury, 2020).

2.2. Informática sanitaria y medios sociales

Los medios sociales y, más específicamente, las redes sociales, se han convertido en una fuente de información y de datos muy importante para poder ser utilizada dentro del ámbito de la informática sanitaria. Se puede decir que la informática sanitaria forma parte de diseñar, desarrollar y aplicar distintas innovaciones basadas en las tecnologías de la información para resolver problemas relacionados con la salud pública y con los servicios sanitarios en general (Gamache et al. 2018). Gracias a esta rama científica interdisciplinar es posible realizar investigaciones complejas con la finalidad de gestionar la información para mejorar la eficiencia y reducir los costes en la asistencia sanitaria (Mei, 2021). La informática sanitaria incluye la ciencia de la información, la informática y la asistencia sanitaria.

La investigación relacionada con la sanidad que utiliza los medios sociales cobra cada vez más importante y puede dividirse en dos grandes ámbitos. El primero se refiere a la vigilancia y predicción en tiempo real de enfermedades como la gripe. Las redes sociales ofrecen un acceso rentable a mensajes localizados geográficamente que contienen debates sobre diferentes temas. Así, es posible investigar las discusiones de los usuarios sobre temas de salud y síntomas de enfermedades (Silk et al. 2019). Basándose en el análisis de la dinámica de las discusiones sobre los síntomas concentradas en determinadas ubicaciones geográficas, se puede identificar o predecir un brote de enfermedad. El segundo ámbito de la investigación de los medios sociales utiliza los datos para obtener una comprensión profunda desde la perspectiva de la paciencia en relación con diferentes condiciones de salud. Los medios sociales son, por tanto, una herramienta rentable para los estudios observacionales y pueden utilizarse para investigar lo que piensan los pacientes sobre diferentes temas sanitarios, como los antibióticos o la vacunación (Latkin et al. 2021; Raftopoulos et al. 2021).

Como se ha señalado anteriormente, Twitter es una fuente de datos popular para la investigación relacionada con la salud. Los expertos identifican tres áreas principales de la investigación sanitaria que utilizan datos de Twitter (Zhang & Ahmed, 2019). En el primer campo de investigación, la minería de datos de las redes sociales generalmente contribuye al conocimiento existente y ayuda a predecir eventos futuros (Zhang & Ahmed, 2019). Esto incluye la vigilancia sindrómica para predecir epidemias de enfermedades estacionales, como la gripe (López et al. 2020; Zeng et al. 2021); la farmacovigilancia (Audeh et al. 2020) y una investigación temática (Mehmet et al. 2020). Los estudios temáticos ayudan a revelar la comprensión del público sobre cuestiones importantes, como la pandemia de COVID-19 (Mheidly & Fares, 2020). Entre los estudios temáticos, el cáncer es el paradigma de investigación más ampliamente enfocado que investiga el contenido compartido y examina el comportamiento de comunicación de los usuarios y las organizaciones sanitarias (Zhang & Ahmed, 2019).

La segunda área de los estudios de informática sanitaria basados en Twitter se refiere a

la naturaleza del intercambio de información en términos de características de los usuarios y de su comportamiento informativo. Esto incluye investigaciones sobre la demografía de los usuarios; la estructura de la red; la frecuencia de la comunicación; los hashtags utilizados y la geografía. La mayoría de los estudios en este ámbito abarcan diversos temas relacionados con el cáncer (Zhang & Ahmed, 2019).

Por último, en el contexto de los medios sociales, la informática sanitaria basada en Twitter incluye estudios que utilizan los datos de los medios sociales para facilitar la comprensión del impacto de los contenidos compartidos (Zhang & Ahmed, 2019). Por ejemplo, Twitter puede utilizarse como un canal de apoyo emocional (L. Liu & Woo, 2021) o como una herramienta para difundir la concienciación e involucrar a la audiencia en la campaña, como el "Día mundial de las enfermedades raras" (Weder et al. 2021). Al analizar el impacto de los contenidos compartidos, es posible revelar cómo sensibilizar con éxito y proporcionar apoyo informativo a través de los medios sociales.

En resumen, los medios sociales son una de las fuentes de datos que pueden servir a los fines de la investigación informática sanitaria. Utilizando los datos de los medios sociales, los investigadores pueden abordar tareas como análisis de comportamiento, análisis de sentimientos, análisis de tendencias y difusión de información. Las tareas de investigación señaladas se refieren generalmente al análisis de contenido que forma parte de las investigaciones de procesamiento del lenguaje natural. Éstas pueden abordarse mediante métodos de aprendizaje automático, como la clasificación, y se analizan en la siguiente sección.

2.3. Métodos de clasificación e informática sanitaria

La investigación propuesta utiliza métodos de aprendizaje automático supervisado para estudiar el amplio debate en las redes sociales sobre la obesidad. El aprendizaje automático supervisado es uno de los métodos más utilizados en el aprendizaje automático debido a su eficiencia (Géron, 2017). El aprendizaje supervisado se utiliza para predecir un determinado resultado basado en una entrada dada mediante la construcción de un par de entrada y

salida. El método tiene como objetivo construir un modelo viable que luego puede utilizarse para realizar predicciones precisas utilizando nuevos datos (Géron, 2017).

La clasificación y la regresión son dos tipos principales de problemas de aprendizaje automático supervisado (Géron, 2017). En la tarea de regresión, el objetivo es predecir un número real. Se puede distinguir de las tareas de clasificación por la continuidad en la salida. En la clasificación, no hay cuestión de grado y el objetivo es predecir una etiqueta de clase de la lista con posibilidades definidas (Géron, 2017). La clasificación es, por tanto, un aprendizaje automático supervisado, ya que utiliza un conjunto de datos de entrenamiento etiquetado para entrenar a los clasificadores. Un conjunto de datos de entrenamiento etiquetado es, por tanto, una condición para la creación de clasificadores. En la minería de textos, una tarea clave es cómo cuantificar los datos textuales. Para ello, se utiliza la ingeniería de rasgos que permite presentar las relaciones entre las palabras representándolas como tokens.

Se ha afirmado que la clasificación es un método eficiente para categorizar grandes conjuntos de datos (Oussous et al. 2018). Permite estudiar los datos textuales y puede proporcionar una visión conceptual mediante la asignación de categorías predeterminadas a los documentos textuales. Este enfoque aporta ventajas muy interesantes para la investigación. La más importante es que generalmente muestra una mayor precisión de las predicciones en comparación con el trabajo humano manual y ahorra considerablemente tiempo (Geirhos et al. 2020; Van Der Walt & Eloff, 2018; Youyou et al. 2015). Para las tareas de clasificación de la informática sanitaria se pueden utilizar diversos métodos. Entre ellos se encuentran Naïve Bayes, Suppor Vector Machine (SVM, o máquinas de vectores de soporte), Random Forest (RF, o bosques aleatorios), Decision Trees (DT, o árboles de decisión), Gradient Boosting Trees (GBT o árboles de refuerzo de gradiente), Gradient Boosted Regression Trees (GBRT, o árboles de regresión con refuerzo de gradiente) y Logistic Regression (LG, o regresión logística).

En el ámbito de la informática sanitaria, se utilizó el algoritmo Naïve Bayes para analizar los datos de las redes sociales y clasificar las publicaciones y predecir los acontecimientos.

Este algoritmo se aplicó para investigar datos de Twitter y predecir brotes de enfermedades, como el dengue y el zika (A.Jabbar Alkubaisi et al. 2018). Otra investigación utilizó Naïve Bayes para probar el clasificar cuatro estados o condiciones de salud, a saber, la gripe, la depresión, el embarazo y los trastornos alimentarios, y dos lugares, Portugal y España (Prieto et al. 2014).

El algoritmo SVM ha demostrado ser una herramienta eficaz para elaborar pronósticos en relación con los brotes de enfermedades. Un estudio de Kesorn y colegas mostró el uso de las SVM para pronosticar brotes de fiebre hemorrágica del dengue a partir de los datos sobre los mosquitos Aedes aegypti que propagan esta enfermedad (Kesorn et al. 2015). Al introducir la función de base radial en el algoritmo clásico, los investigadores lograron una precisión de predicción del 88,37 %. Otros estudios recientes sobre la pandemia de COVID-19 también han mostrado los buenos resultados obtenidos mediante el algoritmo SVM (Dixit et al. 2021; Singh et al. 2020).

Los investigadores aplicaron un clasificador GBRT para desarrollar un modelo de predicción para los usuarios que tuitean sobre los cigarrillos electrónicos (Kim et al. 2017). Se recopilaron 11,5 millones de tuits de 2,6 millones de usuarios en el periodo de noviembre de 2014 y octubre de 2016 utilizando un conjunto de 158 palabras clave. Se analizó una muestra aleatoria de usuarios y sus tuits para desarrollar un protocolo de clasificación de tipos de usuarios aplicando un enfoque de teoría fundamentada. En consecuencia, se clasificaron manualmente 4.897 usuarios utilizando el protocolo desarrollado. A continuación, se obtuvieron quince características de metadatos para cada clase etiquetada a partir de la información del perfil del usuario. A partir del contenido lingüístico y de comportamiento, se obtuvieron 58 características. La ingeniería de características se basa en la matriz de frecuencia de términos y frecuencia inversa de documentos (TF-IDF). Como resultado, el modelo de clasificación mostró una precisión media del 83,3 % para cinco tipos de usuarios.

Los investigadores también pueden aplicar una combinación de algoritmos de clasificación para comparar su rendimiento y luego seleccionar el clasificador más eficaz. Por ejemplo, se

aplicaron los algoritmos Naïve Bayes, Naïve Bayes Multinomial, SVM y árboles de decisión para investigar las tendencias de seguridad alimentaria en Corea a partir de los datos de Twitter(Yeom, 2014). Para el propósito del estudio, se reunieron 14 millones de tweets en 2014. Para estudiar los tuits, se aplicó el Hidden Markov Model (HMM) POS Tagger Work como parte del analizador morfológico Hannanum. Como resultado de la prueba, el clasificador Naive Bayes Multinomial indicó el mejor rendimiento.

Además, un estudio de las redes sociales de salud en línea aplicó la máquina de vectores de apoyo, el bosque aleatorio, el árbol de decisiones, el árbol de aumento gradual y la regresión logística para investigar y predecir las necesidades de los pacientes en relación con el apoyo a la información (Choi et al. 2017). La investigación analizó los datos de 184 usuarios e identificó las principales características que contribuyen a un modelo de clasificación exitoso. Como consecuencia, permite predecir cuatro escenarios, como el apoyo emocional, los hechos médicos, la información basada en la experiencia y la información no convencional. El árbol Gradient Boosting indicó el mejor rendimiento para tres escenarios. En relación con la predicción del apoyo emocional, Gradient Boosting Tree ocupó el segundo lugar después de Support Vector Machine.

Profundizando en los estudios que se han realizado sobre los datos obtenidos de las redes sociales y las TCA, es posible encontrar investigaciones relacionadas con la detección de personas a favor y en contra de algún TCA (Fettach & Benhiba, 2019; Lewis & Arbuthnott, 2012; Oksanen et al. 2015), detección de tuits informativos y no informativos (Viguria et al. 2020; Zhou et al. 2020), e incluso la detección de comunidades de personas que padecen un TCA (Fettach & Benhiba, 2019; Wang et al. 2018). Sin embargo, aún existe la posibilidad de crear modelos predictivos a partir de los datos de Twitter que aporten información interesante sobre la TCA.

Por tanto, los estudios sobre salud basados en datos de medios sociales tienden a aplicar diversos algoritmos y a compararlos, obteniendo diferentes resultados en cada caso particular. Este estudio pretende clasificar temas y, por lo tanto, requiere algoritmos aptos para la

clasificación binaria. En este estudio se han aplicado algunos algoritmos mencionados como bosques aleatorios y máquinas de vectores de soporte, además de otras redes neuronales más complejas como las LSTM y Bi-LSTM, más concretamente los modelos BERT.

Capítulo 3

Metodología de la experimentación

En la sección 3.1 de este capítulo se muestra un flujo de trabajo de la metodología utilizada en este estudio. En la sección 3.2, se describe cómo fue el proceso de recopilación de los mensajes de Twitter sobre los TCA. A continuación, en la sección 3.3, se presentan las tareas de preprocesamiento de datos y etiquetado que fueron realizadas. En la sección 3.4 se presentan los diferentes métodos de clasificación utilizados, incluyendo los antecedentes de Random Forest (RF), Recurrent Neural Network (RNN), Bi-LSTM y de los modelos BERT. Seguidamente, en la sección 3.5, se muestra la configuración utilizada para los distintos métodos de clasificación.

3.1. Flujo de trabajo

En la figura 3.1 se pueden apreciar las diferentes etapas realizadas a lo largo de la presente investigación hasta obtener los modelos de clasificación que dan solución a nuestro problema.

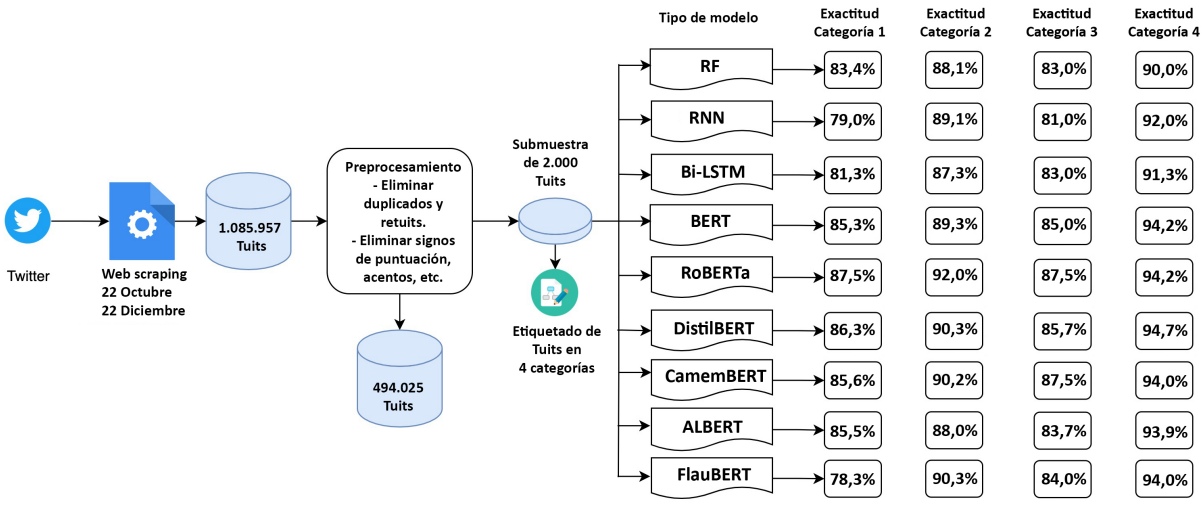

Figura 3.1: *Flujo de trabajo realizado.*

Tal y como se aprecia en la figura anterior, la realización de este proyecto ha supuesto la obtención de los tuits mediante el uso de la API Streaming de Twitter desde el 22 de octubre de 2020 hasta el 22 de diciembre de 2020, el preprocesamiento de los datos, la creación de una submuestra de 2.000 tuits que se etiquetaron manualmente en cuatro categorizaciones diferentes y la creación, entrenamiento y prueba de los diferentes modelos de aprendizaje automático y aprendizaje profundo utilizados en este proyecto.

3.2. Recopilación de datos sobre TCA

Para recopilar los tuits se utilizó una herramienta llamada T-Hoarder (Congosto et al. 2017). Esta herramienta permite seleccionar si se quieren obtener tuits relacionados con determinadas palabras clave o de un usuario concreto, vía API streaming de Twitter. Gracias a su capacidad de obtener datos en directo, fue posible obtener tuits que posteriormente

pudieron ser eliminados. Esto supuso una ventaja sobre todo debido a que la temática de los tuits recopilados, en algunos casos, infringía las normas de la red social y esto nos permitió disponer de una muestra de tuits más grande.

T-hoarder permitió obtener los tuits con la información que se muestra en la tabla 3.1:

La configuración para recopilar los tuits fue la siguiente:

- Se recopilaron todos los tuits que contenían las siguientes palabras clave: *anorexia, anorexic, dietary disorders, inappetence, feeding disorder, food problem, binge eating, anorectic, eating disorders, bulimia, food issues, loss of appetite, food issue, food hater, eat healthier, disturbed eating habits, abnormal eating habits, abnormal eating habit, binge-vomit syndrome, bingeing, bulimarexia, anorexic skinny, eating healthy*. Estas palabras fueron divididas en tres experimentos de T-Hoarder haciendo uso de tres cuentas de Twitter diferentes, lo que permitió obtener un mayor número de tuits sin exceder el límite de uso que tiene la propia plataforma de Twitter.

- Se utilizaron términos en inglés debido a que el número de tuits generados en este idioma es más numeroso que en español dentro de la temática de estudio. No obstante, sería posible recopilar los datos en español y generar los modelos predictivos.

- Se etiquetó manualmente un subconjunto de los tuits recopilados clasificándolos en cuatro categorías diferentes:

    - Categorización 1:
        - Tuits escritos por una persona que padece trastornos alimentarios.
        - Tuits escritos por personas sin trastornos alimentarios
    - Categorización 2:
        - Tuits que fomentan el padecer un TCA.
        - Tuits que no fomentan el padecer un TCA.
    - Categorización 3:

Tabla 3.1: *Descripción de los campos obtenidos mediante T-Hoarder.*

Campo	Descripción
id.tweet	Identificador único de cada tuit. Es un número creciente que está asignando Twitter secuencialmente a cada mensaje.
date	Fecha y hora GMT del tuit.
author	Nombre de usuario del autor del tuit.
text	Texto del tuit.
app	Aplicación desde la que se ha publicado el tuit.
id.author	Identificador del autor. Es un número creciente que se asigna a los usuarios de Twitter a medida que se dan de alta.
followers	Número de seguidores en el momento de la publicación.
following	Número de usuarios seguidos en el momento de la publicación.
statuses	Número de tuits publicados anteriormente.
location	Ubicación declarada en el perfil del usuario.
urls	Enlace si el tuit contiene una URL, en caso contrario, se almacena el valor None.
geolocation	Coordenadas si el tuit está geolocalizado, en caso contrario se almacena None.
name	Nombre proporcionado por el usuario.
description	Descripción del usuario
url_media	URL si el tuit contiene información multimedia, en caso contrario, se almacena el valor None.
type_media	Tipo de información multimedia (foto, vídeo,...), en caso de no existir, su valor es None.
quoted	Tuit original que ha sido citado
relation	RT, respuesta, cita o None.
replied_id	Tuit respondido.
user.replied	Usuario respondido.
retweeted_id	Id del tuit respondido.
user_retweeted	Usuario retuiteado.
quoted_id	Id del tuit citado.
user.quoted	Usuario citado.
first.HT	Primer hashtag encontrado en el tuit.
lang	Idioma del tuit.
create_at	Fecha de creación de la cuenta de usuario de Twitter
verified	Información sobre si la cuenta está verificada o no.
avatar	Avatar del usuario.
link	Enlace al tuit.

    - Tuits de carácter informativo.
    - Tuits de opinión.
  - Categorización 4:
    - Tuits de carácter científico.
    - Tuits de carácter no científico.

- Es posible ver un ejemplo de cada una de las categorizaciones en la tabla 3.2.
- Para realizar esta categorización ha sido necesario acceder uno a uno a cada perfil de usuario que ha enviado un tuit de este tipo.

Tabla 3.2: *Ejemplos de tuits categorizados.*

Categoría	Tema	Tuit
Tipo 1	Escrito por alguien que padece TCA	i was stressed and ate a whole bowl of pasta, where's my badge for being the worst anorexic #edtwt
Tipo 1	Escrito por alguien que no padece TCA	Is your #teenager not eating or eating a lot less than normal? She might be suffering from #anorexia. We can help; please come see us https://t.co/GfStM1IVGz #weightloss #losingweight https://t.co/z5NK0tjNIt
Tipo 2	Fomenta los TCA	Currently feeling like the best anorexic #EDtwt https://t.co/1BZPMs8bGU
Tipo 2	No fomenta los TCA	Higher-calorie diets could lead to a speedier recovery in patients with anorexia nervosa, study shows https://t.co/mipX3nrhHN #mentalhealth #diet #anorexia
Tipo 3	Informativo	#AnorexiaNervosa – A Father and Daughter Perspective - Highlights from RCPsychIC 2019 #EatingDisorders #mentalhealth https://t.co/iq3GH5ce6C
Tipo 3	De opinión	Binge eating makes me sad :(#eatingdisorder #bingeeating https://t.co/0jjf7YrVyc
Tipo 4	Científico	The problem extends to Food and Drug Administration and National Institutes of Health datasets used in a recent study appearing in Reproductive Toxicology. #ai #technology #BigData #ML https://t.co/DFvh6gNA38
Tipo 4	No científico	Do not waste time thinking about what you could have done differently. Keep your eyes on the road ahead and do it differently now. #anorexia #eatingdisorder #recovery #nevergiveup #alwayskeepfighting https://t.co/YalYzclBDM

3.3. Preprocesamiento de los datos y etiquetado

Para realizar el preprocesamiento de los datos se programó en lenguaje *Python* en su versión 3.6. Se describe el procedimiento a continuación:

1. **Carga de datos en un *dataframe*:** Se cargaron los datos de los documentos obtenidos mediante T-Hoarder que son nombrados de la forma 'streaming_experimento_número.txt'. T-Hoarder genera un fichero hasta que este llega a tener un tamaño de 100MB. Por ello, se obtuvieron 4 ficheros para el conjunto de datos 1, 4 ficheros para el conjunto de datos 2 y dos ficheros para el conjunto de datos 3. Los conjuntos de datos que se recopilaron fueron los siguientes:

   - Conjunto 1: *anorexia, anorexic, dietary disorders, inappetence, feeding disorder, food problem, binge eating, anorectic*
   - Conjunto 2: *eating disorders, bulimia, food issues, loss of appetite, food issue, food hater, eat healthier, disturbed eating habits, abnormal eating habits, abnormal eating habit*
   - Conjunto 3: *binge-vomit syndrome, bingeing, bulimarexia, anorexic skinny, eating healthy*

   Esta división fue realizada planteando los dos primeros conjunto como aquellos que contenían los términos más comunes, y un tercer conjunto con elementos menos comunes.

2. **Eliminación de saltos de línea y tabulaciones**: Algunos datos, como la ubicación, el nombre y la biografía, pueden contener saltos de línea o tabulaciones. Para evitar conflictos con los delimitadores, se filtraron los tabuladores y los saltos de línea de estos datos. Para ello se generó una función a la cual se le pasaba como parámetro los distintos *dataframes* y preprocesaba los datos.

3. **Concatenación de los *dataframes*:** Tras el preprocesamiento de los dataframes, se concatenaron todos en un único *dataframe*.

4. **Reducción de tamaño del *dataframe*:** Para poder trabajar de forma más ágil con el *dataframe* de datos, se calculó el uso de memoria del mismo y se optimizó realizando tres tareas fundamentales: (i) convertir las columnas númericas a números, convertir las fechas al formato *datetime* y convertir el resto de objetos en categorías. Estos pasos ayudaron a reducir el *dataframe* de 2,7GB a 1,1GB.

5. **Eliminación de retuits:** Se realizó un cruce entre las columnas *id.tweet* y *retweeted_id* para eliminar todos los tuits que son retuits de tuits ya existentes en el *dataframe*.

6. **Eliminación de tuits duplicados:** se eliminaron los posibles tuits repetidos ya que, inicialmente, se unificaron conjuntos de datos que podrían contener tuits comunes y, además, filtramos solo aquellos tuits escritos en inglés.

7. **Etiquetado de una muestra de 2.000 tuits:** Tras el preprocesamiento de los datos realizado explicado en los pasos anteriores, se redujo la muestra de tuits de 1.085.957 a 494.025. De estos 494.025, un se etiquetó manualmente una muestra de 2.000 tuits categorizados en las 4 categorías expuestas anteriormente: (i) tuits escritos por personas que padecen, o no, TCA, (ii) tuits que fomentan, o no, el padecer un TCA, (iii) tuits de carácter informativo, o de opinión y (iv) tuits de carácter científico, o no.

8. **Preprocesamiento de la muestra final:** A esta muestra final se le realizó un preprocesamiento mediante el cual se eliminaron las *stopwords* así como otros signos puntuación y símbolos que dificultaban la aplicación de técnicas de aprendizaje automático. Las *stopwords* (palabras vacías) son aquellas palabras que carecen de significado por sí solas y que modifican o acompañan a otras, por ejemplo, artículos, pronombres, adverbios, preposiciones o algunos verbos.

El código de esto procedimiento se encuentra en el **Anexo A**.

3.4. Métodos de clasificación utilizados

Se aplicaron distintos métodos de clasificación de aprendizaje automático y de aprendizaje profundo. Estos métodos de clasificación fueron: Random Forest (RF), Recurrent Neural Networks (RNN), redes Bi-LSTM y modelos BERT preentrenados. En relación a los métodos RNN, Bi-LSTM y BERT, fueron elegidos porque, según la revisión de la literatura mencionada en el capítulo 2 de esta memoria, parecen ser los modelos más prometedores en cuestión de PLN. Con respecto a RF, fueron elegidos para tener una referencia de un modelo más sencillo y así poder comparar los resultados obtenidos y el coste computacional en relación a los demás.

Para la aplicación de los modelos BERT se hizo uso de *ClassificationModel* de la biblioteca *simpletransformers* (disponible en https://github.com/ThilinaRajapakse/simpletransformers) para clasificar el modelo. Se utilizó este modelo porque es adecuado para clasificar texto en forma binaria dentro de esta biblioteca. Se han entrenado seis redes neuronales utilizando diferentes tipos de modelos BERT (BERT, RoBERTa (Y. Liu et al. 2019), DistilBERT (Sanh et al. 2020), CamemBERT (Martin et al. 2020), ALBERT (Lan et al. 2020) y FlauBERT (Le et al. 2020)), todos ellos utilizando modelos preentrenados.

La biblioteca *scikit-learn* se utilizó para dividir el conjunto de datos utilizando *train_ test_ split* y también para obtener la métrica *f1 score* y la métrica de exactitud (*accuracy*) de los modelos de clasificación generados.

Además, para RF se utilizó una validación cruzada con una $k = 10$, mientras que para las redes neuronales se realizaron 5 iteraciones diferentes y, posteriormente, se obtuvo la media para las métricas *f1 score* y *accuracy*.

3.4.1. Random Forest, RNN y Bi-LSTM

Random Forest

Random Forest (RF), o bosque aleatorio, es un algoritmo de aprendizaje automático supervisado. El "bosque" que construye es un conjunto de árboles de decisión, normalmente entrenados con el método "bagging". La idea general de este método es que una combinación de modelos de aprendizaje aumenta el resultado global.

En pocas palabras: RF construye múltiples árboles de decisión y los fusiona para obtener una predicción más precisa y estable.

Una gran ventaja de RF es que puede utilizarse tanto para problemas de clasificación como de regresión, que son la mayoría de los sistemas actuales de aprendizaje automático.

RF tiene casi los mismos hiperparámetros que un árbol de decisión. Afortunadamente, no hay necesidad de combinar un árbol de decisión con otros métodos porque se puede utilizar fácilmente la clase de clasificador de RF. Con RF, también se pueden tratar las tareas de regresión utilizando el regresor del algoritmo.

Además, RF añade aleatoriedad adicional al modelo, mientras crecen los árboles. En lugar de buscar la característica más importante al dividir un nodo, busca la mejor característica entre un subconjunto aleatorio de características. Esto da lugar a una amplia diversidad que generalmente da lugar a un mejor modelo.

Por lo tanto, en el bosque aleatorio, el algoritmo sólo tiene en cuenta un subconjunto aleatorio de características para dividir un nodo. Incluso puede hacer que los árboles sean más aleatorios utilizando adicionalmente umbrales aleatorios para cada característica en lugar de buscar los mejores umbrales posibles (como hace un árbol de decisión normal).

RNN

Una Recurrent Neural Network (RNN), o red neuronal recurrente, es una clase de redes neuronales artificiales en las que las conexiones entre los nodos forman un gráfico dirigido a lo largo de una secuencia temporal. Esto le permite mostrar un comportamiento dinámico

temporal. Derivadas de las redes neuronales feedforward, las RNN pueden utilizar su estado interno (memoria) para procesar secuencias de entrada de longitud variable, lo que las hace aplicables a tareas como el reconocimiento de escritura manuscrita no segmentada y conectada o el reconocimiento del habla.

El término "red neuronal recurrente" se utiliza indistintamente para referirse a dos amplias clases de redes con una estructura general similar, donde una es de impulso finito y la otra de impulso infinito. Una red recurrente de impulso finito es un grafo acíclico dirigido que puede ser desenrollado y sustituido por una red neuronal estrictamente feedforward, mientras que una red recurrente de impulso infinito es un grafo cíclico dirigido que no puede ser desenrollado.

Tanto las redes recurrentes de impulso finito como las de impulso infinito pueden tener estados adicionales almacenados, y el almacenamiento puede estar bajo control directo de la red neuronal. El almacenamiento también puede ser sustituido por otra red o grafo, si éste incorpora retrasos temporales o tiene bucles de retroalimentación. Estos estados controlados se denominan estado controlado o memoria controlada, y forman parte de las redes Long Short-Term Memory (LSTM), o redes de memoria a corto plazo, y de las unidades recurrentes controladas. También se denominan Feedforward Neural Network (FNN), o redes neuronales de retroalimentación.

Bi-LSTM

Las LSTM bidireccionales son una extensión de las LSTM tradicionales que pueden mejorar el rendimiento del modelo en problemas de clasificación de secuencias.

En los problemas en los que todos los pasos de tiempo de la secuencia de entrada están disponibles, las LSTM bidireccionales entrenan dos en lugar de una LSTM en la secuencia de entrada. El primero en la secuencia de entrada tal cual y el segundo en una copia invertida de la secuencia de entrada. Esto puede proporcionar un contexto adicional a la red y dar lugar a un aprendizaje más rápido e incluso más completo del problema.

3.4.2. Antecedentes sobre modelos BERT

El modelo Bidirectional Encoder Representations from Transformers (BERT) ha causado sensación en la comunidad de aprendizaje automático al mostrar los últimos resultados en una variedad de tareas de PLN (Procesamiento del Lenguaje Natural), incluyendo la respuesta a preguntas (SQuAD v1.1), la inferencia del lenguaje natural (MNLI), etc (Devlin et al. 2019).

Una de las principales innovaciones técnicas añadidas en los modelos BERT es la aplicación del entrenamiento bidireccional de *Transformer*, un popular modelo de atención, al modelado del lenguaje. Anteriormente, las secuencias de texto se examinaban de izquierda a derecha o con un entrenamiento combinado de izquierda a derecha y de derecha a izquierda. Los resultados obtenidos por (Devlin et al. 2019) muestran que los modelos lingüísticos entrenados bidireccionalmente tienen un sentido más profundo del contexto y del flujo del lenguaje que los modelos lingüísticos unidireccionales. Además, los investigadores destacan una técnica novedosa llamada *Masked LM* (MLM). Esta técnica permite el entrenamiento bidireccional de modelos que antes eran imposibles de entrenar con el método tradicional.

En el campo de la visión por ordenador, se ha demostrado repetidamente la importancia del aprendizaje por transferencia. Este aprendizaje se produce tras el preentrenamiento de una red neuronal en una tarea conocida, por ejemplo, ImageNet, y el posterior procesamiento para mejorar el modelo. Utilizando una red neuronal entrenada como base, se consigue un nuevo modelo con un objetivo específico.

Algunos investigadores de PLN han demostrado que este enfoque también puede utilizarse en tareas de PLN. Utilizando una red neuronal preentrenada, es posible incrustar palabras que luego se utilizan como características de los modelos de PLN.

En este trabajo proponemos crear cuatro modelos predictivos relacionados con los trastornos de la conducta alimentaria utilizando un conjunto de datos propio obtenido de la red social Twitter. Para ello, proponemos el uso de un enfoque basado en el modelo BERT para clasificar tuits en base a las cuatro categorías mencionadas anteriormente en este do-

cumento. Se estudiarán diferentes modelos BERT y se determinará cuál es el que obtiene los mejores resultados.

3.4.3. Cómo trabajan los modelos BERT

Los modelos BERT se basan en un mecanismo de atención que es capaz de aprender las relaciones contextuales entre palabras o subpalabras dentro de un texto. Este mecanismo se conoce como *Transformer* (Vaswani et al. 2017) y cuenta con dos mecanismos: uno capaz de leer un texto de entrada conocido como clasificador y otro que realiza una predicción para la tarea conocido como decodificador. En este caso, para los modelos BERT, de los 2 que ofrece *Transformer*, solo es necesario el mecanismo de codificación.

La principal ventaja de Transformer sobre los modelos direccionales es que, a diferencia de éstos, el codificador lee toda la secuencia de palabras a la vez, de izquierda a derecha y de derecha a izquierda, no en dos secuencias diferentes. Por eso BERT se considera bidireccional, aunque sería más exacto decir que es no direccional. Gracias a esta característica, los modelos pueden aprender el contexto de una palabra basándose en todas las palabras que la rodean, no sólo en las que están a ambos lados.

Al entrenar modelos lingüísticos, uno de los retos es definir un objetivo de predicción claro. Los modelos que predicen la siguiente palabra de una secuencia con un enfoque direccional están inherentemente limitados al aprendizaje del contexto. En estos casos, BERT hace uso de dos enfoques conocidos como LM enmascarado (MLM) (Kushilevitz et al. 2020) y Predicción de la siguiente frase (NSP) (Shi & Demberg, 2019).

3.5. Configuración de los métodos de clasificación utilizados

Para poder aplicar los métodos de clasificación se utilizó un cuaderno Jupyter sobre Python 3.6 mediante el cual se utilizaron las librerías Tensorflow y Pytorch. Fue necesario utilizar ambas librerías debido a que, por el momento, las redes BERT solo pueden ser

generadas a través de Pytorch, mientras que Tensorflow es una de las librerías más utilizadas para generar modelos RF, RNN o Bi-LSTM. La división de los datos consistió en un 70 % de datos de entrenamiento y un 30 % de datos de validación.

A continuación se detalla información sobre los métodos de clasificación utilizados y sus hiperparámetros:

- **Random Forest (RF):** Para aplicar RF a las cuatro categorías se hizo uso de la librería *sklearn* y se importó *RandomForestClassifier*. Tras realizar algunas pruebas, se encontró que el entrenamiento de modelos RF con 1.000 estimadores proporcionó los mejores resultados.

- **Recurrent Neural Networks (RNN):** Para entrenar modelos mediante redes neuronales recurrentes se utilizó **sklearn** y la librería **keras** de Tensorflow. Al tratarse de categorizaciones binarias, la función de activación utilizada fue *sigmoid*.

- **Bi-LSTM personalizado:** Para la creación, entrenamiento y validación de los modelos Bi-LSTM, se hizo uso de las librerías **sklearn** y Tensorflow.

- **Modelos BERT preentrenados**: Para hacer uso de modelos preentrenados BERT se utilizó principalmente la librería *simpletransformers* (disponible en https://github.com/ThilinaRajapakse/simpletransformers) importando la clase *ClasificacionModel* junto con la librería Pytorch. Se implementaron seis *ClassificationModels* con los argumentos mencionados y habiendo elegido seis tipos de modelos diferentes haciendo uso de seis modelos preentrenados, que se explican en la sección **??**.

Capítulo 4

Experimentos y resultados

4.1. Resultados de la recopilación de datos y preprocesamiento

Los datos se recogieron entre el 20 de octubre de 2020 y el 26 de diciembre de 2020, consiguiendo un total de 1.085.957 tuits.

Se realizó un preprocesamiento de los datos que incluyó las siguientes tareas:

- Eliminación de todos los tuits que tenían el mismo mensaje.
- Eliminación de aquellos retuits de los que no disponíamos del tuit original.
- Eliminación de los tuits que no fueron recogidos correctamente.
- Se han eliminado las menciones.
- Se han eliminado los acentos en el texto, así como los signos de puntuación.
- También se han eliminado las *stopwords*.

Tras este procedimiento, se obtuvieron 494.025 tuits válidos. Este conjunto de datos está disponible en un repositorio abierto en el siguiente enlace: https://www.kaggle.com/jabenitez88/eating-disorders-tweets.

Las palabras más utilizadas en los tuits son las que se observan en la figura 4.1.

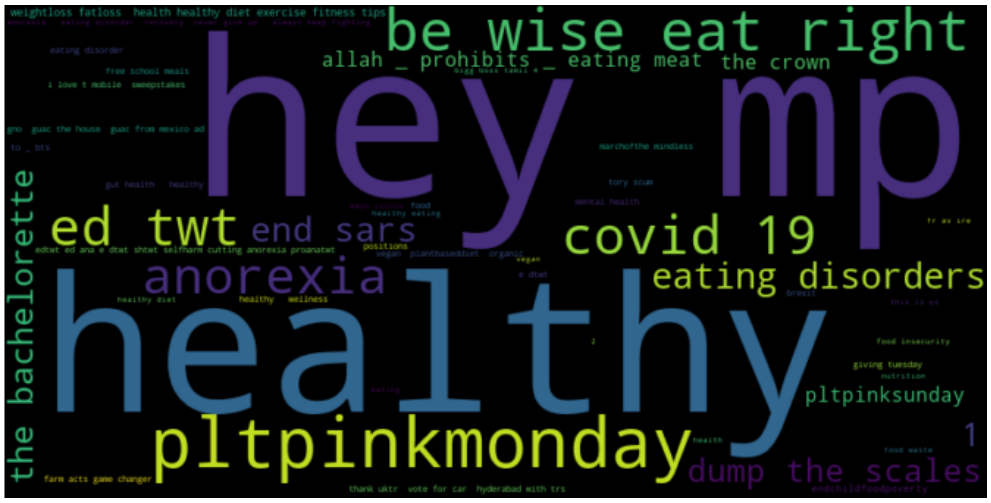

Figura 4.1: *Palabras más repetidas en los 494.025 tuits válidos recopilados.*

Tras este preprocesamiento, se creó un subconjunto de 2.000 tuits y se etiquetaron en las cuatro categorizaciones mencionadas anteriormente. Las palabras más utilizadas fueron las que se observan en la figura 4.2

Figura 4.2: *Palabras más repetidas en los 2.000 tuits categorizados.*

La relación de tuits recopilados en cada categoría fue la siguiente:

- Categorización 1:
  - El 48,5 % de tuits fueron escritos por una persona que padece trastornos alimen-

tarios.

  - El 51,5 % de tuits fueron escritos por personas sin trastornos alimentarios.

- Categorización 2:

  - El 23,4 % de tuits fomentan el padecer un TCA.
  - El 76,6 % de tuits no fomentan el padecer un TCA.

- Categorización 3:

  - El 38,4 % de tuits fueron de carácter informativo.
  - El 61,6 % de tuits fueron de opinión.

- Categorización 4:

  - El 24,5 % de tuits fueron de carácter científico.
  - El 75,5 % de tuits fueron de carácter no científico.

Es posible apreciar esta distribución en un gráfico de columnas en la figura 4.3.

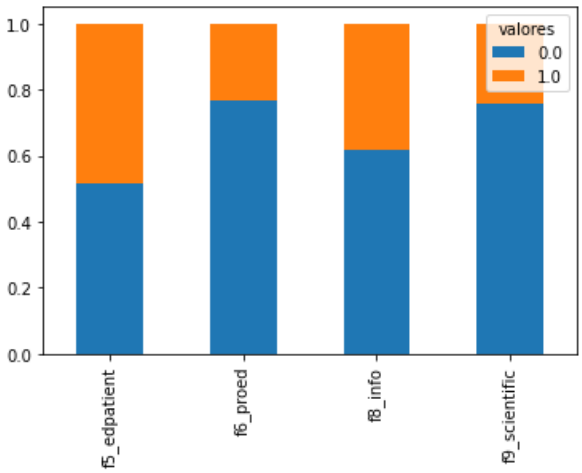

Figura 4.3: *Representación de porcentajes de tuits pertenecientes a cada una de las cuatro categorías.*

4.2. Resultados

A continuación se muestran los resultados obtenidos en las métricas f1, exactitud (acc) y tiempo de ejecución tras entrenar los distintos modelos con las cuatro categorizaciones diferentes en las tablas 4.1, 4.2, 4.3 y 4.4.

Tabla 4.1: *Resultados obtenidos con los diferentes modelos para la clasificación de la categorización 1.*

Tipo de modelo	Modelo preentrenado	f1	acc	tiempo
Random Forest		0,831	83,4 %	94,3s
RNN		0,770	79,0 %	152,1s
Bi-LSTM		0,780	81,3 %	163,2s
BERT	bert-based-multilingual-cased	0,848	85,3 %	1257,4s
RoBERTa	roberta-base	0,868	87,5 %	1116,2s
DistilBERT	distilbert-base-cased	0,856	86,3 %	1343,3s
CamemBERT	camembert-base	0,838	85,6 %	1472,3s
ALBERT	albert-base-v1	0,849	85,5 %	1372,7s
FlauBERT	flaubert_base_cased	0,757	78,3 %	1203,9s

Tabla 4.2: *Resultados obtenidos con los diferentes modelos para la clasificación de la categorización 2.*

Tipo de modelo	Modelo preentrenado	f1	acc	tiempo
Random Forest		0,711	88,1 %	99,2s
RNN		0,800	89,1 %	163,1s
Bi-LSTM		0,780	87,3 %	175,3s
BERT	bert-based-multilingual-cased	0,746	89,3 %	1232,1s
RoBERTa	roberta-base	0,818	92,0 %	1158,8s
DistilBERT	distilbert-base-cased	0,780	90,3 %	1327,8s
CamemBERT	camembert-base	0,763	90,2 %	1457,5s
ALBERT	albert-base-v1	0,714	88,0 %	1352,3s
FlauBERT	flaubert_base_cased	0,784	90,3 %	1207,1s

Tabla 4.3: *Resultados obtenidos con los diferentes modelos para la clasificación de la categorización 3.*

Tipo de modelo	Modelo preentrenado	f1	acc	tiempo
Random Forest		0,790	83,0 %	95,3s
RNN		0,780	81,0 %	151,5s
Bi-LSTM		0,793	83,0 %	164,8s
BERT	bert-based-multilingual-cased	0,811	85,0 %	1292,7s
RoBERTa	roberta-base	0,840	87,5 %	1142,5s
DistilBERT	distilbert-base-cased	0,809	85,7 %	1332,0s
CamemBERT	camembert-base	0,847	87,5 %	1462,0s
ALBERT	albert-base-v1	0,791	83,7 %	1331,3s
FlauBERT	flaubert_base_cased	0,797	84,0 %	1202,1s

Tabla 4.4: *Resultados obtenidos con los diferentes modelos para la clasificación de la categorización 4.*

Tipo de modelo	Modelo preentrenado	f1	acc	tiempo
Random Forest		0,83	90,0 %	93,4s
RNN		0,860	92,0 %	148,4s
Bi-LSTM		0,853	91,3 %	154,3s
BERT	bert-based-multilingual-cased	0,888	94,2 %	1272,4s
RoBERTa	roberta-base	0,880	94,2 %	1149,1s
DistilBERT	distilbert-base-cased	0,890	94,7 %	1328,5s
CamemBERT	camembert-base	0,885	94,0 %	1403,6s
ALBERT	albert-base-v1	0,876	93,9 %	1302,9s
FlauBERT	flaubert_base_cased	0,875	94,0 %	1227,1s

Los porcentajes de mejora entre el mejor modelo BERT y el mejor modelo de entre RF, RNN y Bi-LSTM son los siguientes:

- **Categorización 1:** RoBERTa (87,5 %) vs RF (83,4 %) = 4,92 % de mejora.
- **Categorización 2:** RoBERTa (92,0 %) vs RNN (89,1 %) = 3,25 % de mejora.
- **Categorización 3:** RoBERTa (87,5 %) vs RF (83,0 %) = 5,42 % de mejora.
- **Categorización 4:** DistilBERT (94,7 %) vs RNN (92,0 %) = 2,94 % de mejora.

Capítulo 5

Discusión y conclusiones

Este capítulo presenta las respuestas a las preguntas de investigación planteadas y analiza los resultados de la investigación. En primer lugar, se discute sobre si ha sido posible conseguir modelos de aprendizaje automático o aprendizaje profundo para clasificar de forma precisa los tuits sobre TCA. En segundo lugar, se exponen cuáles han sido los modelos más precisos. A continuación, aclara las implicaciones de los resultados. Por último, expone las limitaciones de la investigación.

Mediante esta investigación se han recogido una serie de mensajes a través de la red social Twitter, todos ellos relacionados con los trastornos alimentarios. Se ha logrado cumplir con el objetivo inicial de la investigación presentada en esta memoria consistente en generar modelos predictivos capaces de clasificar los tweets determinando si pertenecían a una de las siguiente cuatro categorías: (i) tuits escritos por personas que padecen TCA, o no, (ii) tuits que fomentan el padecer un TCA, o no, (iii) tuits informativos o de opinión y (iv) tuits con carácter científico, o no.

Después de preprocesar los datos y de que se etiquetasen manualmente un subconjunto de 2.000 tuits, dicho subconjunto fue posteriormente preprocesado en profundidad. Una vez realizada esta tarea, los datos se dividieron en conjuntos de entrenamiento (70 %) y de prueba o validación (30 %). Se entrenaron modelos de tipo Random Forest (RF), Recurrent Neural Network (RNN), redes Bi-LSTM personalizadas y, además, 6 modelos BERT preentrenados diferentes.

De los modelos entrenados según las diferentes categorías se observaron los siguientes resultados:

- **Categorización 1:** Esta clasificación de tuits consistía en detectar tuits escritos por personas que han padecido un TCA o no. El modelo que mayor exactitud obtuvo fue el modelo RoBERTa con un 87,5 % de exactitud. Todos los modelos BERT, salvo FlauBERT, superaron el 85,0 %. Sin embargo, aplicando bosques aleatorios se obtuvo una exactitud de un 83,4 %.

- **Categorización 2:** Esta clasificación permitía detectar tuits que fomentaban padecer un TCA o no. El modelo que mayor exactitud obtuvo fue el modelo RoBERTa con un 92,0 % de exactitud. Todos los modelos preentrenados BERT, superaron o igualaron el 88,0 %. Sin embargo, aplicando bosques aleatorios se obtuvo una exactitud de un 88,1 % y aplicando RNN se obtuvo una exactitud de un 89,1 %.

- **Categorización 3:** Esta clasificación permitía detectar tuits de tipo informativo o de opinión. El modelo que mayor exactitud obtuvo fue el modelo RoBERTa y CamemBERT con un 87,5 % de exactitud. Todos los modelos BERT, salvo AlBERT y FlauBERT, superaron el 85,0 %. Sin embargo, aplicando bosques aleatorios se obtuvo una exactitud de un 83,0 %.

- **Categorización 4:** El modelo que mayor exactitud obtuvo fue el modelo DistilBERT con un 94,7 % de exactitud. Todos los modelos BERT, salvo AlBERT, superaron o igualaron el 94,0 %. Sin embargo, aplicando bosques aleatorios se obtuvo una exactitud de un 90,0 % y aplicando RNN un 92,0 %.

Se han utilizado datos obtenidos mediante medios sociales, más concretamente de la plataforma Twitter. Esta recopilación fue posible gracias una herramienta llamada T-hoarder que hace uso de API streaming de Twitter (Congosto et al. 2017). Como resultado, se obtuvieron 1.085.953 tuits que fueron preprocesados, quedando un total de 494.025 tuits. De este conjunto se etiquetó un subconjunto de 2.000.

Analizando el coste computacional que generan los modelos BERT preentrenados frente al modelo más sencillo de bosques aleatorios o a una RNN simple nos haría pensar si realmente es necesario, para el caso de estudio en particular, hacer uso de modelos BERT o no. La exactitud obtenida en los diferentes modelos, para las cuatro categorizaciones o clasificaciones diferentes, es bastante alta incluso en los modelos más simples.

En este estudio se ha podido observar que, a pesar de contar con una muestra de datos de tan solo 2.000 Tweets, el rendimiento de los diferentes modelos de clasificación aplicados ha sido bastante bueno, lo que resulta muy prometedor para el desarrollo de futuros modelos predictivos dentro del campo de los TCA y otras enfermedades mentales, entre otros.

Ha sido posible comparar y mejorar los resultados obtenidos a través de esos modelos predictivos en base a los resultados obtenidos.

Tras todos los análisis realizados, es posible destacar que los académicos podrían utilizar los clasificadores para extraer mensajes útiles para su posterior análisis en el ámbito de los TCA. Los profesionales se beneficiarían. También es posible destacar que, aunque el modelo que mejor exactitud tiene siempre ha sido uno de los modelos BERT preentrenado, el coste computacional que ello supone frente a otros modelos más simples puede resultar excesivo. La diferencia entre los porcentajes de exactitud del mejor modelo BERT frente al mejor modelo de entre los tres más simples (RF, RNN y Bi-LSTM), no exceden el 5,42 %.

Esta investigación tiene varias limitaciones, entre ellas (i) se limita a una plataforma de medios sociales, (ii) algunas categorizaciones no están balancedas, lo que puede provocar un sesgo en los modelos generados, (iii) el conjunto de entrenamiento ha sido suficiente, pero puede ser mayor, mejorando notablemente los resultados en un entorno real. Por todo ello, se proponen algunas posibles mejoras futuras a continuación:

- Aumentar el conjunto de datos de entrenamiento y validación contando con un mayor número de mensajes etiquetados.
- Aplicar técnicas de PLN que hagan uso de ontologías mediante las cuales sea posible plantear automatizaciones y razonamientos lógicos incluyendo la creación de sistemas

expertos que hagan uso de todo ello.

- Integrar los modelos predictivos en un proyecto de desarrollo real como, por ejemplo, un bot de Twitter capaz de detectar si los tuits están siendo escritos por personas que padezcan algún TCA.
- Investigar en mayor profundidad los efectos de los TCA en la salud, especialmente desde una perspectiva pública.
- Hacer uso de otras plataformas de medios sociales disponibles.

.

Capítulo 6

Planificación y estimación de costes

6.1. Planificación del trabajo

La planificación temporal de este trabajo ha supuesto un total de 359 horas de trabajo por parte del estudiante, lo que suponen un total de 44 días laborables de 8 horas de trabajo diarias. La estimación del tiempo que supone cada una de las tareas realizadas no excede los 13 días de duración. En la tabla 6.1 se detallan las diferentes tareas y la duración aproximada que han supuesto cada una de ellas.

Tabla 6.1: *Estimación de trabajo y coste.*

Tipo de tarea	Duración
Revisión bibliográfica de la literatura	6 días
Instalación del entorno de recopilación de tuits	5 días
Instalación de programas	2 días
Tratamiento de los datos en crudo	8 días
Etiquetado de la submuestra de 2.000 tuits	10 días
Programación de los algoritmos de aprendizaje automático	13 días

6.2. Estimación de costes

6.2.1. Especificación de recursos

En los proyectos de ingeniería y ciencia de datos, cuando se habla de recursos, se hace referencia, principalmente, al personal necesario para realizar el trabajo.

Para poder realizar este proyecto se cuenta con varios perfiles, aunque, en este caso, todos ellos están representados por una sola persona:

Tabla 6.2: *Salario medio según el rol de los trabajadores.*

Recursos	Salario/año	Salario/mes	Salario/día	Salario/hora
Ingeniero en informática	26.404,00 €	2.200,33 €	110,02 €	13,75 €
Ingeniero de datos	36.001,00 €	3.000,08 €	150,00 €	18,75 €
Científico de datos	32.937,00 €	2.744,75 €	137,24 €	17,15 €

Para estimar el salario de cada rol de trabajo se ha utilizado la plataforma web indeed. Esta plataforma permite obtener la media salarial para cada puesto de trabajo basándose en los datos disponibles para España durante el año 2020. Los salarios por mes, día y hora se han calculado mediante las siguientes fórmulas:

$$\text{Salario/mes} = \text{Salario anual} \ / \ 12$$

$$\text{Salario/día} = \text{Salario mes} \ / \ 20$$

$$\text{Salario/hora} = \text{Salario día} \ / \ 8$$

6.2.2. Asignación de recursos

Cada tarea a realizar en el proyecto es asignada a los recursos humanos disponibles. En esta ocasión, al realizar todas las tareas una única persona, no es posible realizar tareas en paralelo, con lo que las jornadas de 8 horas de trabajo diarias se reparten de forma proporcional entre los distintos recursos humanos involucrados para cada una de las tareas.

En la tabla 6.3 es posible ver la estimación del trabajo a realizar según los diferentes roles junto con el coste:

Tabla 6.3: *Estimación de trabajo y coste.*

Recurso	Trabajo (horas)	Coste (€)
Ingeniero en Informática	38	522,50
Ingeniero de datos	147	2.756,25
Científico de datos	174	2.984,10
Total	**359**	**6.262,85**

Tal y como se aprecia en la tabla anterior, el proyecto conlleva un total de **359h** de trabajo equivalente a **44 días** laborables. Esto supone un coste total de **6.262,85 €** en lo que a recursos humanos se refiere.

En relación a los costes materiales, se ha realizado una división consistente en: costes hardware (equipamiento informático) y costes software (programas informáticos). Para cada uno de estos costes se detalla el coste unitario, el número de unidades y el total. En la tabla 6.4 se pueden observar los costes hardware:

Tabla 6.4: *Estimación de costes de Hardware.*

Nombre	Coste unitario	Unidades	Coste total
Ordenador sobremesa	2.450,00 €	1	2.450,00 €
Ratón	12,00 €	1	12,00 €
Monitor	165,00 €	2	330,00 €
Servidor VPS	400,00 €	1	400,00 €
Total hardware	**-**	**5**	**3.192,00 €**

En la tabla 6.4 se pueden observar los costes hardware y en la tabla 6.5 los costes software.

Tabla 6.5: *Estimación de costes de Software.*

Nombre	Coste unitario	Unidades	Coste total
Licencia de Microsoft Office	579,00 €	1	579,00 €
Licencia de Microsoft Project	849,00 €	1	849,00 €
Licencia de Windows 10	116,98 €	1	116,98 €
Total Software	**-**	**3**	**1.544,98 €**

Destacar que, a nivel de hardware, para realizar el preprocesamiento de los tuits y aplicar los diferentes modelos de clasificación se utilizó en un ordenador con las siguientes características: CPU Intel(R) Core(TM) i7-9700K a 3,60GHZ, 32,0GB de RAM y una tarjeta gráfica NVIDIA GeForce RTX 2080.

6.2.3. Presupuesto

Un presupuesto detallado debe contar con los recursos humanos y con los recursos materiales que se deben emplear. Se desglosan los costes en la tabla 6.6.

Para realizar una estimación del presupuesto, se ha de tener en cuenta tanto los recursos humanos, como los materiales empleados. El coste total estimado se muestra desglosado en la siguiente tabla:

Tabla 6.6: *Presupuesto detallado para realizar.*

Tipo de coste	Coste	Porcentaje
Recursos humanos	6.262,85 €	56,94 %
Hardware	3.192,00 €	29,02 %
Software	1.544,98 €	14,05 %
Total	**10.999,83 €**	**100 %**

En la figura 6.1 es posible ver un gráfico circular en el que se detallan los porcentajes de gastos según el tipo de recurso utilizado.

El coste total estimado del proyecto es de **10.999,83€**, donde el mayor gasto se refleja en los recursos humanos (56,94 %), seguido del hardware (29,02 %) y, en último lugar, el software (14,05 %).

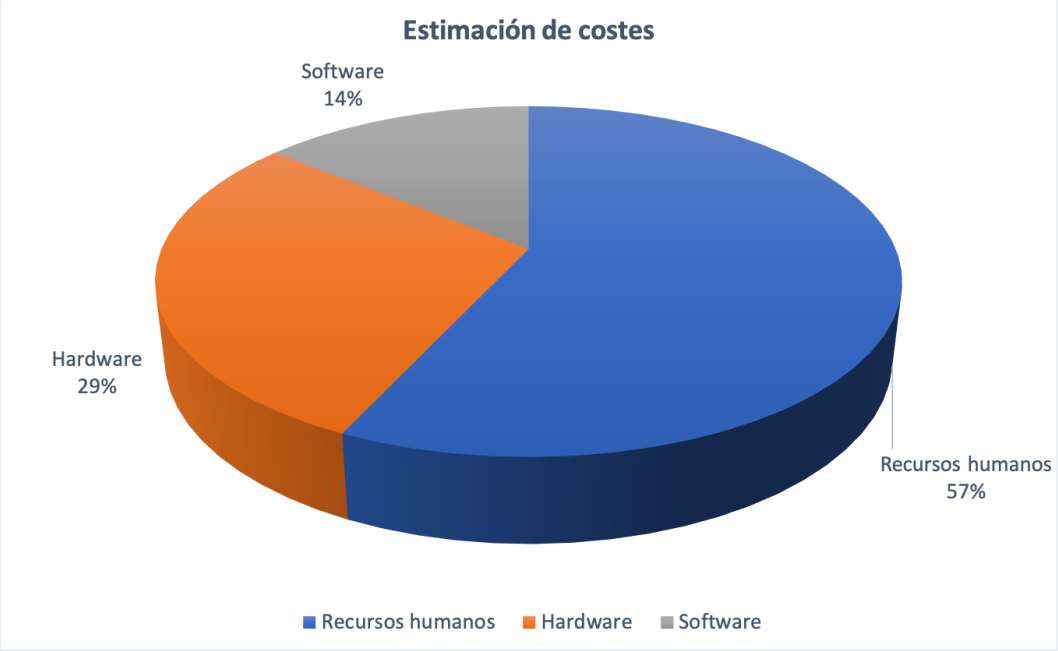

Figura 6.1: *Porcentaje estimado de costes según el tipo de recurso.*

Capítulo 7

Aportaciones científicas realizadas

Este trabajo de fin de máster ha dado lugar a una comunicación de congreso presentada el de 9 de junio de 2021 en el congreso "34th IEEE CBMS International Symposium on Computer-Based Medical Systems" de forma virtual.

La comunicación fue titulada "BERT Model-Based Approach For Detecting Categories of Tweets in the Field of Eating Disorders (ED)" y los autores del mismo fueron:

- José Alberto Benítez-Andrades
- José Manuel Alija-Pérez
- Isaías García-Rodríguez
- Carmen Benavides
- Hector Alaiz Moreton
- Rafael Pastor-Vargas
- María Teresa García-Ordás

Se presenta la comunicación como **Anexo B** en este documento.

Referencias

Ackland, R. (2009). Social Network Services as Data Sources and Platforms for e-Researching Social Networks. *Social Science Computer Review*, *27*(4), 481-492. https://doi.org/10.1177/0894439309332291

A.Jabbar Alkubaisi, G. A., Kamaruddin, S. S. & Husni, H. (2018). Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers. *Computer and Information Science*, *11*(1), 52. https://doi.org/10.5539/cis.v11n1p52

Arigo, D., Pagoto, S., Carter-Harris, L., Lillie, S. E. & Nebeker, C. (2018). Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery. *DIGITAL HEALTH*, *4*, 205520761877175. https://doi.org/10.1177/2055207618771757

Audeh, B., Bellet, F., Beyens, M.-N., Lillo-Le Louët, A. & Bousquet, C. (2020). Use of Social Media for Pharmacovigilance Activities: Key Findings and Recommendations from the Vigi4Med Project. *Drug Safety*, *43*(9), 835-851. https://doi.org/10.1007/s40264-020-00951-2

Bauer, G. R. & Lizotte, D. J. (2021). Artificial Intelligence, Intersectionality, and the Future of Public Health. *American Journal of Public Health*, *111*(1), 98-100. https://doi.org/10.2105/AJPH.2020.306006

Blank, G. (2017). The Digital Divide Among Twitter Users and Its Implications for Social Research. *Social Science Computer Review*, *35*(6), 679-697. https://doi.org/10.1177/0894439316671698

Boyd, D. M. & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210-230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Carceller-Maicas, N. (2016). Youth, health and social networks: Instagram as a research tool for health communication. *Metode*, *0*(6), 227-233. https://doi.org/10.7203/metode.6.6555

Chancellor, S. & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review [Number: 1 Publisher: Nature Publishing Group]. *npj Digital Medicine*, *3*(1), 1-11. https://doi.org/10.1038/s41746-020-0233-7

Choi, M.-J., Kim, S.-H., Lee, S., Kwon, B. C., Yi, J. S., Choo, J. & Huh, J. (2017). Toward Predicting Social Support Needs in Online Health Social Networks [Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada]. *Journal of Medical Internet Research*, *19*(8), e7660. https://doi.org/10.2196/jmir.7660

Congosto, M., Basanta-Val, P. & Sanchez-Fernandez, L. (2017). T-Hoarder: A framework to process Twitter data streams. *Journal of Network and Computer Applications*, *83*, 28-39. https://doi.org/10.1016/j.jnca.2017.01.029

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171-4186. https://doi.org/10.18653/v1/N19-1423

Dixit, A., Mani, A. & Bansal, R. (2021). CoV2-Detect-Net: Design of COVID-19 prediction model based on hybrid DE-PSO with SVM using chest X-ray images. *Information Sciences*. https://doi.org/https://doi.org/10.1016/j.ins.2021.03.062

Edo-Osagie, O., Iglesia, B. D. L., Lake, I. & Edeghere, O. (2020). A scoping review of the use of Twitter for public health research. *Computers in Biology and Medicine*, *122*, 103770. https://doi.org/https://doi.org/10.1016/j.compbiomed.2020.103770

Fettach, Y. & Benhiba, L. (2019). Pro-eating disorders and pro-recovery communities on reddit: Text and network comparative analyses. *ACM International Conference Proceeding Series*, 277-286. https://doi.org/10.1145/3366030.3366058

Fiumara., G., Celesti., A., Galletta., A., Carnevale., L. & Villari., M. Applying Artificial Intelligence in Healthcare Social Networks to Identity Critical Issues in Patients' Posts. En: *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5: AI4Health,* INSTICC. SciTePress, 2018, 680-687. ISBN: 978-989-758-281-3. https://doi.org/10.5220/0006750606800687.

Galmiche, M., Déchelotte, P., Lambert, G. & Tavolacci, M. P. (2019). Prevalence of eating disorders over the 2000-2018 period: A systematic literature review. https://doi.org/10.1093/ajcn/nqy342

Gamache, R., Kharrazi, H. & Weiner, J. P. (2018). Public and Population Health Informatics: The Bridging of Big Data to Benefit Communities. *Yearbook of Medical Informatics*, *27*(1), 199-206. https://doi.org/10.1055/s-0038-1667081

Geirhos, R., Meding, K. & Wichmann, F. A. (2020). Beyond accuracy: quantifying trial-by-trial behaviour of CNNs and humans by measuring error consistency [arXiv: 2006.16736]. *arXiv:2006.16736 [cs, q-bio]*. Consultado el 5 de junio de 2021, desde http://arxiv.org/abs/2006.16736
Comment: NeurIPS 2020 camera ready

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow concepts, tools, and techniques to build intelligentsystems* (1.ª ed.). http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20%5C&path=ASIN/1491962291

Griffen, T. C., Naumann, E. & Hildebrandt, T. (2018). Mirror exposure therapy for body image disturbances and eating disorders: A review. https://doi.org/10.1016/j.cpr.2018.08.006

Harris, D. L. & Carr, A. T. (2001). Prevalence of concern about physical appearance in the general population. *British Journal of Plastic Surgery*, *54*(3), 223-226. https://doi.org/10.1054/bjps.2001.3550

Hoek, H. W. (2016). Review of the worldwide epidemiology of eating disorders. *Current Opinion in Psychiatry*, *29*(6), 336-339. https://doi.org/10.1097/YCO.0000000000000282

Home - Global Nutrition Report. (s.f.). Consultado el 8 de diciembre de 2020, desde https://globalnutritionreport.org/

Izquierdo, A., Plessow, F., Becker, K. R., Mancuso, C. J., Slattery, M., Murray, H. B., Hartmann, A. S., Misra, M., Lawson, E. A., Eddy, K. T. & Thomas, J. J. (2019). Implicit attitudes toward dieting and thinness distinguish fat-phobic and non-fat-phobic anorexia nervosa from avoidant/restrictive food intake disorder in adolescents. *International Journal of Eating Disorders*, *52*(4), 419-427. https://doi.org/10.1002/eat.22981

Jordan, S. E., Hovet, S. E., Fung, I. C.-H., Liang, H., Fu, K.-W. & Tse, Z. T. H. (2019). Using Twitter for Public Health Surveillance from Monitoring and Prediction to Public Response. *Data*, *4*(1). https://doi.org/10.3390/data4010006

Kesorn, K., Ongruk, P., Chompoosri, J., Phumee, A., Thavara, U., Tawatsin, A. & Siriyasatien, P. (2015). Morbidity Rate Prediction of Dengue Hemorrhagic Fever (DHF) Using the Support Vector Machine and the Aedes aegypti Infection Rate in Similar Climates and Geographical Areas [Publisher: Public Library of Science]. *PLOS ONE*, *10*(5), e0125049. https://doi.org/10.1371/journal.pone.0125049

Kim, A., Miano, T., Chew, R., Eggers, M. & Nonnemaker, J. (2017). Classification of Twitter Users Who Tweet About E-Cigarettes [Company: JMIR Public Health and Surveillance Distributor: JMIR Public Health and Surveillance Institution: JMIR Public Health and Surveillance Label: JMIR Public Health and Surveillance Publisher: JMIR Publications Inc., Toronto, Canada]. *JMIR Public Health and Surveillance*, *3*(3), e8060. https://doi.org/10.2196/publichealth.8060

Kushilevitz, G., Markovitch, S. & Goldberg, Y. (2020). A Two-Stage Masked LM Method for Term Set Expansion. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6829-6835. https://doi.org/10.18653/v1/2020.acl-main.610

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P. & Soricut, R. (2020). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations.

Latkin, C., Dayton, L. A., Yi, G., Konstantopoulos, A., Park, J., Maulsby, C. & Kong, X. (2021). COVID-19 vaccine intentions in the United States, a social-ecological framework. *Vaccine*, *39*(16), 2288-2294. https://doi.org/https://doi.org/10.1016/j.vaccine.2021.02.058

Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L. & Schwab, D. (2020). FlauBERT: Unsupervised Language Model Pre-training for French. *Proceedings of the 12th Language Resources and Evaluation Conference*, 2479-2490. https://www.aclweb.org/anthology/2020.lrec-1.302

Lewis, S. P. & Arbuthnott, A. E. (2012). Searching for Thinspiration: The Nature of Internet Searches for Pro-Eating Disorder Websites. *Cyberpsychology, Behavior, and Social Networking*, *15*(4), 200-204. https://doi.org/10.1089/cyber.2011.0453

Liu, L. & Woo, B. K. P. (2021). Twitter as a Mental Health Support System for Students and Professionals in the Medical Field [Company: JMIR Medical Education Distributor:

JMIR Medical Education Institution: JMIR Medical Education Label: JMIR Medical Education Publisher: JMIR Publications Inc., Toronto, Canada]. *JMIR Medical Education*, *7*(1), e17598. https://doi.org/10.2196/17598

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.

Lobstein, T., Baur, L. & Uauy, R. (2004). Obesity in children and young people: A crisis in public health. https://doi.org/10.1111/j.1467-789x.2004.00133.x

Lobstein, T., Jackson-Leach, R., Moodie, M. L., Hall, K. D., Gortmaker, S. L., Swinburn, B. A., James, W. P. T., Wang, Y. & McPherson, K. (2015). Child and adolescent obesity: Part of a bigger picture. https://doi.org/10.1016/S0140-6736(14)61746-3

López, L., Fernández, M., Gómez, A. & Giovanini, L. (2020). An influenza epidemic model with dynamic social networks of agents with individual behaviour. *Ecological Complexity*, *41*, 100810. https://doi.org/10.1016/j.ecocom.2020.100810

Lopez-Castroman, J., Moulahi, B., Azé, J., Bringay, S., Deninotti, J., Guillaume, S. & Baca-Garcia, E. (2020). Mining social networks to improve suicide prevention: A scoping review. *Journal of Neuroscience Research*, *98*(4), 616-625. https://doi.org/10.1002/jnr.24404

Lou, C. & Tse, C. H. (2020). Which model looks most like me? Explicating the impact of body image advertisements on female consumer well-being and consumption behaviour across brand categories. *International Journal of Advertising*, 1-27. https://doi.org/10.1080/02650487.2020.1822059
doi: 10.1080/02650487.2020.1822059

Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D. & Sagot, B. (2020). CamemBERT: a Tasty French Language Model. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7203-7219. https://doi.org/10.18653/v1/2020.acl-main.645

Martínez-Rojas, M., Pardo-Ferreira, M. d. C. & Rubio-Romero, J. C. (2018). Twitter as a tool for the management and analysis of emergency situations: A systematic literature review. *International Journal of Information Management*, *43*, 196-208. https://doi.org/10.1016/j.ijinfomgt.2018.07.008

Martín-Rojas, R., García-Morales, V. J., Garrido-Moreno, A. & Salmador-Sánchez, M. P. (2021). Social Media Use and the Challenge of Complexity: Evidence from the Technology Sector. *Journal of Business Research*, *129*, 621-640. https://doi.org/https://doi.org/10.1016/j.jbusres.2019.12.026

Mehmet, M., Roberts, R. & Nayeem, T. (2020). Using digital and social media for health promotion: A social marketing approach for addressing co-morbid physical and mental health [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajr.12589]. *Australian Journal of Rural Health*, *28*(2), 149-158. https://doi.org/https://doi.org/10.1111/ajr.12589

Mei, R. (2021). Health Informatics and Healthcare Delivery: From the Cost-effectiveness Perspective. *2021 7th International Conference on Information Management (ICIM)*, 62-65. https://doi.org/10.1109/ICIM52229.2021.9417045

Mheidly, N. & Fares, J. (2020). Leveraging media and health communication strategies to overcome the COVID-19 infodemic. *Journal of Public Health Policy*, *41*(4), 410-420. https://doi.org/10.1057/s41271-020-00247-w

Musacchio, N., Giancaterini, A., Guaita, G., Ozzello, A., Pellegrini, M. A., Ponzani, P., Russo, G. T., Zilich, R. & de Micheli, A. (2020). Artificial intelligence and big data in diabetes care: A position statement of the Italian association of medical diabetologists. https://doi.org/10.2196/16922

Obar, J. A. & Wildman, S. (2015). Social Media Definition and the Governance Challenge: An Introduction to the Special Issue. *Internal Communications & Organizational Behavior eJournal*.

Oksanen, A., Garcia, D., Sirola, A., Näsi, M., Kaakinen, M., Keipi, T. & Räsänen, P. (2015). Pro-anorexia and anti-pro-anorexia videos on YouTube: Sentiment analysis of user responses. *Journal of Medical Internet Research*, *17*(11), e256. https://doi.org/10.2196/jmir.5007

Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A. & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, *30*(4), 431-448. https://doi.org/10.1016/j.jksuci.2017.06.001

Prieto, V. M., Matos, S., Álvarez, M., Cacheda, F. & Oliveira, J. L. (2014). Twitter: A Good Place to Detect Health Conditions [Publisher: Public Library of Science]. *PLOS ONE*, *9*(1), e86191. https://doi.org/10.1371/journal.pone.0086191

Raftopoulos, V., Iordanou, S., Katsapi, A., Dedoukou, X. & Maltezou, H. C. (2021). A comparative online survey on the intention to get COVID-19 vaccine between Greek and Cypriot healthcare personnel: is the country a predictor? [PMID: 33754953]. *Human Vaccines & Immunotherapeutics*, *0*(0), 1-8. https://doi.org/10.1080/21645515.2021.1896907

Sanh, V., Debut, L., Chaumond, J. & Wolf, T. (2020). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.

Senín-Calderón, C., Gálvez-González, J., Perona-Garcelán, S., Camacho, C. & Rodríguez-Testal, J. F. (2020). Dysmorphic concern and behavioural impairment related to body image in adolescents. *International Journal of Psychology*, *55*(5), 832-841. https://doi.org/10.1002/ijop.12646

Shi, W. & Demberg, V. (2019). Next Sentence Prediction helps Implicit Discourse Relation Classification within and across Domains. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 5790-5796. https://doi.org/10.18653/v1/D19-1586

Silk, M. J., Hodgson, D. J., Rozins, C., Croft, D. P., Delahay, R. J., Boots, M. & McDonald, R. A. (2019). Integrating social behaviour, demography and disease dynamics in network models: applications to disease management in declining wildlife populations [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1781), 20180211. https://doi.org/10.1098/rstb.2018.0211

Singh, V., Poonia, R. C., Kumar, S., Dass, P., Agarwal, P., Bhatnagar, V. & Raja, L. (2020). Prediction of COVID-19 corona virus pandemic based on time series data using sup-

port vector machine. *Journal of Discrete Mathematical Sciences and Cryptography*, *23*(8), 1583-1597. https://doi.org/10.1080/09720529.2020.1784535

Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L. & Merchant, R. M. (2017). Twitter as a tool for health research: A systematic review. https://doi.org/10.2105/AJPH.2016.303512

Sohota, J. (2020). Social media: the good, the bad and the ugly. *BDJ In Practice*, *33*(5), 18-19. https://doi.org/10.1038/s41404-020-0391-y

Statista. (2021). *Most popular social networks worldwide as of January 2021, ranked by number of active users*. Consultado el 6 de mayo de 2021, desde https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Stellefson, M., Paige, S. R., Chaney, B. H. & Chaney, J. D. (2020). Evolving Role of Social Media in Health Promotion: Updated Responsibilities for Health Education Specialists. *International Journal of Environmental Research and Public Health*, *17*(4). https://doi.org/10.3390/ijerph17041153

Timmins, K. A., Green, M. A., Radley, D., Morris, M. A. & Pearce, J. (2018). How has big data contributed to obesity research? A review of the literature. https://doi.org/10.1038/s41366-018-0153-7

Urdapilleta, I., Lahlou, S., Demarchi, S. & Catheline, J. M. (2019). Women with obesity are not as curvy as they think: Consequences on their everyday life behavior. *Frontiers in Psychology*, *10*(AUG). https://doi.org/10.3389/fpsyg.2019.01854

Vaganov, D., Bardina, M. & Guleva, V. (2020). From Generality to Specificity: On Matter of Scale in Social Media Topic Communities. En V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos & J. Teixeira (Eds.), *Computational Science – ICCS 2020* (pp. 305-318). Springer International Publishing.

Van Der Walt, E. & Eloff, J. (2018). Using Machine Learning to Detect Fake Identities: Bots vs Humans [Conference Name: IEEE Access]. *IEEE Access*, *6*, 6540-6549. https://doi.org/10.1109/ACCESS.2018.2796018

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is All you Need. En I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett (Eds.), *Advances in Neural Information Processing Systems*. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Viguria, I., Alvarez-Mon, M. A., Llavero-Valero, M., del Barco, A. A., Ortuño, F. & Alvarez-Mon, M. (2020). Eating disorder awareness campaigns: Thematic and quantitative analysis using twitter. *Journal of Medical Internet Research*, *22*(7), e17626. https://doi.org/10.2196/17626

Wang, T., Brede, M., Ianni, A. & Mentzakis, E. (2018). Social interactions in online eating disorder communities: A network perspective (M. Tang, Ed.). *PLOS ONE*, *13*(7), e0200800. https://doi.org/10.1371/journal.pone.0200800

Weder, F., Krainer, L. & Karmasin, M. (2021). *The Sustainability Communication Reader: A Reflective Compendium* [Google-Books-ID: iUAjEAAAQBAJ]. Springer Nature.

World Health Organization. (2020). Obesity and overweight. Consultado el 8 de diciembre de 2020, desde https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Yeom, H.-N. (2014). Study of Machine-Learning Classifier and Feature Set Selection for Intent Classification of Korean Tweets about Food Safety -Journal of Information Science Theory and Practice | Korea Science [Publisher: Korea Institute of Science and Technology Information]. *Journal of Information Science Theory and Practice*, *2*(3), 29-39.

Youyou, W., Kosinski, M. & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans [Publisher: National Academy of Sciences Section: Social Sciences]. *Proceedings of the National Academy of Sciences*, *112*(4), 1036-1040. https://doi.org/10.1073/pnas.1418680112

Zarrinkalam, F., Piao, G., Faralli, S. & Bagheri, E. (2020). Mining User Interests from Social Media. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 3519-3520. https://doi.org/10.1145/3340531.3412167

Zeng, D., Cao, Z. & Neill, D. B. (2021). Chapter 22 - Artificial intelligence–enabled public health surveillance—from local detection to global epidemic monitoring and control. En L. Xing, M. L. Giger & J. K. Min (Eds.), *Artificial Intelligence in Medicine* (pp. 437-453). Academic Press. https://doi.org/10.1016/B978-0-12-821259-2.00022-3

Zhang, Z. & Ahmed, W. (2019). A comparison of information sharing behaviours across 379 health conditions on Twitter. *International Journal of Public Health*, *64*(3), 431-440. https://doi.org/10.1007/s00038-018-1192-5

Zhou, S., Zhao, Y., Bian, J., Haynos, A. F. & Zhang, R. (2020). Exploring Eating Disorder Topics on Twitter: Machine Learning Approach. *JMIR Medical Informatics*, *8*(10), e18273. https://doi.org/10.2196/18273

Anexo A

Parte del código desarrollado en este trabajo

A.1. Código desarrollado para el preprocesamiento de los datos

- **Código para cargar los datos en un dataframe:**

```
import os
import pandas as pd

data1_0 = pd.read_csv('streaming_ed-data1_0.txt', encoding='utf8',
    sep="\t", header = None, error_bad_lines=False)
```

- **Código para eliminar saltos de línea y tabulaciones:**

```
def pre_df(data, group):
    data.columns =
     ["id_tweet","date","author","text","app","id_user","followers",
    "following","statuses","location","urls","geolocation","name",
    "description", "url_media","type_media","quoted","relation",
    "replied_id","user_replied","retweeted_id","user_retweeted",
    "quoted_id","user_quoted","first_HT", "lang",
    "created_at","verified","avatar","link"]
```

```
    data['app'] = data['app'].astype('str').map(lambda x:
    ↪ x.lstrip('via='))
    data['id_user'] = data['id_user'].astype('str').map(lambda x:
    ↪ x.lstrip('id='))
    data['followers'] = data['followers'].astype('str').map(lambda x:
    ↪ x.lstrip('followers='))
    data['following'] = data['following'].astype('str').map(lambda x:
    ↪ x.lstrip('following='))
    data['statuses'] = data['statuses'].astype('str').map(lambda x:
    ↪ x.lstrip('statuses='))
    data['location'] = data['location'].astype('str').map(lambda x:
    ↪ x.lstrip('loc='))
    data['stream_group'] = group
```

- **Concatenación de los *dataframes*:** Tras el preprocesamiento de los dataframes, se concatenaron todos en un único *dataframe* de la siguiente manera:

```
objs = [data1_0, data1_1, data1_2, data1_3, data2_0, data2_1, data2_2,
↪ data2_3, data3_0, data3_1]

data_tot = pd.concat(
    objs,
    axis=0,
    join="outer",
    ignore_index=True,
    copy=True
)
```

- **Código para reducir el espacio del dataframe:**

```
def usageForType(df):
    for ctype in
    ↪ ['float','float64','int64','int','object','datetime','category']:
        columnType = df.select_dtypes(include=[ctype])
```

```
        meanMemoryUsage = columnType.memory_usage(deep=True).mean()
        meanMemoryUsageMB = meanMemoryUsage / 1024 ** 2
        print("Memory usage for type ",ctype , " : {:0.5f}
        ↪  MB".format(meanMemoryUsageMB))

data_tot_cl.loc[:,'id_tweet'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'id_tweet'], errors='coerce')
data_tot_cl.loc[:,'id_user'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'id_user'], errors='coerce')
data_tot_cl.loc[:,'followers'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'followers'], errors='coerce')
data_tot_cl.loc[:,'following'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'following'], errors='coerce')
data_tot_cl.loc[:,'replied_id'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'replied_id'], errors='coerce')
data_tot_cl.loc[:,'retweeted_id'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'retweeted_id'], errors='coerce')
data_tot_cl.loc[:,'quoted_id'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'quoted_id'], errors='coerce')
data_tot_cl.loc[:,'user_replied'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'user_replied'], errors='coerce')
data_tot_cl.loc[:,'user_retweeted'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'user_retweeted'],
↪  errors='coerce')
data_tot_cl.loc[:,'user_quoted'] =
↪  pd.to_numeric(data_tot_cl.loc[:,'user_quoted'], errors='coerce')

data_tot_cl['date'] = pd.to_datetime(data_tot_cl['date'],
↪  errors='coerce', infer_datetime_format=True)
data_tot_cl['created_at'] = pd.to_datetime(data_tot_cl['created_at'],
↪  errors='coerce', infer_datetime_format=True)

for col in data_tot_cl.columns:
       if "object" in str(data_tot_cl[col].dtype):
            data_tot_cl[col] = data_tot_cl[col].astype('category')
```

- **Eliminación de retuits:** Se realizó un cruce entre las columnas *id.tweet* y *retweeted_id* para eliminar todos los tuits que son retuits de tuits ya existentes en el *dataframe*.

```
id_tweets = set(data_tot_cl['id_tweet'].unique())
data_tot_cl["origtweetishere"] =
    data_tot_cl['retweeted_id'].map(lambda x : True if x in id_tweets
    else False)
```

- **Eliminación de tuits duplicados:** se eliminaron los posibles tuits repetidos ya que, inicialmente, se unificaron conjuntos de datos que podrían contener tuits comunes y, además, filtramos solo aquellos tuits escritos en inglés.

```
data_tot_cl = data_tot_cl.drop_duplicates()
data_tot_cl = data_tot_cl.loc[~data_tot_cl['text'].duplicated()]
data_tot_cl = data_tot_cl.loc[~data_tot_cl['id_tweet'].duplicated()]
data_tot_cl = data_tot_cl.query('lang == "en"')
```

- **Preprocesamiento de la muestra final eliminando stopwords:**

```
punctuations = "¡!#$%&'()*+,-./:;<=>¿?@[\]^_`{|}~"

def read_txt(filename):
    list = []
    with open(filename, 'r', encoding='utf-8') as f:
        data = f.readlines()
        for line in data:
            list.append(str(line).replace('\n', ''))
    return list

stopwords = read_txt('english_stopwords.txt')

```

```
13 stemmer = SnowballStemmer('english')
14
15
16 def clean_accents(tweet):
17     tweet = re.sub(r"[àáâãäå]", "a", tweet)
18     tweet = re.sub(r"ç", "c", tweet)
19     tweet = re.sub(r"[èéêë]", "e", tweet)
20     tweet = re.sub(r"[ìíîï]", "i", tweet)
21     tweet = re.sub(r"[òóôõö]", "o", tweet)
22     tweet = re.sub(r"[ùúûü]", "u", tweet)
23     tweet = re.sub(r"[ýÿ]", "y", tweet)
24
25     return tweet
26
27 def clean_tweet(tweet, stem = False):
28     tweet = tweet.lower().strip()
29     tweet = re.sub(r'https?:\/\/\S+', '', tweet)
30     tweet = re.sub(r'http?:\/\/\S+', '', tweet)
31     tweet = re.sub(r'www?:\/\/\S+', '', tweet)
32     tweet = re.sub(r'\s([@#][\w_-]+)', "", tweet)
33     tweet = re.sub(r"\n", " ", tweet)
34     tweet = clean_accents(tweet)
35     tweet =
       ↪  re.sub(r"\b(a*ha+h[ha]*|o?l+o+l+[ol]*|x+d+[x*d*]*|a*ja+[j+a+]+)\b",
       ↪  "<risas>", tweet)
36     for symbol in punctuations:
37         tweet = tweet.replace(symbol, "")
38     tokens = []
39     for token in tweet.strip().split():
40         if token not in punctuations and token not in stopwords:
41             if stem:
42                 tokens.append(stemmer.stem(token))
43             else:
44                 tokens.append(token)
45     return " ".join(tokens)
46
```

```
tweets1['text_cleaned'] = tweets['text_orig'].apply(lambda s :
    clean_tweet(s))
print(tweets1['text_cleaned'].head(5))
```

A.2. Código generado para crear los distintos modelos predictivos

- **RandomForest:**

```
punctuations = "¡!#$%&'()*+,-./:;<=>¿?@[\]^_`{|}~"

X1_train, X1_test, y1_train, y1_test = train_test_split(X, Y1,
    test_size=0.3, random_state=42)
X2_train, X2_test, y2_train, y2_test = train_test_split(X, Y2,
    test_size=0.3, random_state=42)
X3_train, X3_test, y3_train, y3_test = train_test_split(X, Y3,
    test_size=0.3, random_state=42)
X4_train, X4_test, y4_train, y4_test = train_test_split(X, Y4,
    test_size=0.3, random_state=42)

from sklearn.ensemble import RandomForestClassifier

classifier = RandomForestClassifier(n_estimators=1000, random_state=0)
classifier.fit(X1_train, y1_train)

y1_pred = classifier.predict(X1_test)

from sklearn.metrics import classification_report, confusion_matrix,
    accuracy_score

print(confusion_matrix(y1_test,y1_pred))
print(classification_report(y1_test,y1_pred))
print(accuracy_score(y1_test, y1_pred))
```

- **Recurrent Neural Network (RNN):**

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from keras.models import Model
from keras.layers import LSTM, Activation, Dense, Dropout, Input,
    Embedding
from keras.optimizers import RMSprop
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
from keras.utils import to_categorical
from keras.callbacks import EarlyStopping

df = tweets1.copy()
X = df['text_cleaned']
Y1 = df['f5_edpatient']
Y2 = df['f6_proed']
Y3 = df['f8_info']
Y4 = df['f9_scientific']

X1_train, X1_test, y1_train, y1_test = train_test_split(X, Y1,
    test_size=0.3, random_state=42)
X2_train, X2_test, y2_train, y2_test = train_test_split(X, Y2,
    test_size=0.3, random_state=42)
X3_train, X3_test, y3_train, y3_test = train_test_split(X, Y3,
    test_size=0.3, random_state=42)
X4_train, X4_test, y4_train, y4_test = train_test_split(X, Y4,
    test_size=0.3, random_state=42)

max_words = 1000
max_len = 150
```

```
tok = Tokenizer(num_words=max_words)
tok.fit_on_texts(X1_train)
sequences = tok.texts_to_sequences(X1_train)
sequences_matrix = sequence.pad_sequences(sequences,maxlen=max_len)

def RNN():
    inputs = Input(name='inputs',shape=[max_len])
    layer = Embedding(max_words,50,input_length=max_len)(inputs)
    layer = LSTM(100)(layer)
    layer = Dense(256,name='FC1')(layer)
    layer = Activation('relu')(layer)
    layer = Dropout(0.1)(layer)
    layer = Dense(1,name='out_layer')(layer)
    layer = Activation('sigmoid')(layer)
    model = Model(inputs=inputs,outputs=layer)
    return model

X1_train, X1_test, y1_train, y1_test = train_test_split(X, Y3,
↪  test_size=0.3, random_state=42)


max_words = 1000
max_len = 150
tok = Tokenizer(num_words=max_words)
tok.fit_on_texts(X1_train)
sequences = tok.texts_to_sequences(X1_train)
sequences_matrix = sequence.pad_sequences(sequences,maxlen=max_len)
model = RNN()
model.summary()
model.compile(loss='binary_crossentropy',optimizer=RMSprop(),
metrics=['accuracy'])
model.fit(sequences_matrix,y1_train,batch_size=64,epochs=20,
          validation_split=0.2,
          callbacks=[EarlyStopping(monitor='val_loss'
          ,min_delta=0.0001)])
test_sequences = tok.texts_to_sequences(X1_test)
```

```
test_sequences_matrix =
    sequence.pad_sequences(test_sequences,maxlen=max_len)
accr = model.evaluate(test_sequences_matrix,y1_test)
print('Test set\n  Loss: {:0.3f}\n  Accuracy:
    {:0.3f}'.format(accr[0],accr[1]))
matrix = sklearn.metrics.confusion_matrix(y1_test, y1_pred)
matrix
```

- **Bi-LSTM personalizado:**

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
import tensorflow as tf

X1_train, X1_test, y1_train, y1_test = train_test_split(X, Y1,
    test_size=0.3, random_state=42)

VOCAB_SIZE=1000
encoder =
    tf.keras.layers.experimental.preprocessing.TextVectorization(
    max_tokens=VOCAB_SIZE)
encoder.adapt(np.asarray(X1_train))

model = tf.keras.Sequential([
    encoder,
    tf.keras.layers.Embedding(
        input_dim=len(encoder.get_vocabulary()),
        output_dim=64,
        # Use masking to handle the variable sequence lengths
        mask_zero=True),
    tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(64)),
```

```
    tf.keras.layers.Dense(64, activation='relu'),
    tf.keras.layers.Dense(1)
])

model.compile(loss=tf.keras.losses.BinaryCrossentropy(from_logits=True),
              optimizer=tf.keras.optimizers.Adam(2e-4),
              metrics=['accuracy'])

history = model.fit(X1_train,y1_train, epochs=14,
                    validation_data=(X1_test,y1_test),
                    validation_steps=30)

test_loss, test_acc = model.evaluate(X1_test,y1_test)

print('Test Loss: {}'.format(test_loss))
print('Test Accuracy: {}'.format(test_acc))

matrix = sklearn.metrics.confusion_matrix(y1_test, y1_pred)
matrix
```

- **Modelos BERT preentrenados**:

```
import pandas as pd
import numpy as np

import spacy
import nltk
import nltk.data
from nltk.tokenize import word_tokenize
from nltk.stem import SnowballStemmer
import regex as re
import string
from collections import defaultdict

import matplotlib.pyplot as plt
```

```
import seaborn as sns
pd.set_option('display.max_colwidth', None)

import sklearn
from sklearn.model_selection import train_test_split
from sklearn.metrics import f1_score, accuracy_score

from simpletransformers.classification import ClassificationModel


import io


train_args ={"reprocess_input_data": True,
           "fp16":False,
             "evaluate_during_training": False,
             "evaluate_during_training_verbose":False,
             "learning_rate":2e-5,
             "train_batch_size":32,
             "eval_batch_size":32,
           "num_train_epochs": 15, 'overwrite_output_dir': True,
            "evaluation_strategy":'epochs'
            }
#optimizer = torch.optim.SGD(model.parameters(), lr=0.01,
    momentum=0.9)


def f1_multiclass(labels, preds):
    return f1_score(labels, preds, average='micro')

def calcule_f1(df):
    return(df['tp'] / (df['tp'] + 0.5 * (df['fp'] + df['fn'])))

import torch
import gc
from tqdm import tqdm
```

```
48
49
50  dfEval1 = pd.DataFrame()
51
52
53  indexBERT =
    ↪  ['BERT','RoBERTa','DistilBERT','CamemBERT','Albert','FlauBERT']
54
55  train_df1 = pd.DataFrame({ 'text_cleaned': X1_train, 'target':
    ↪  y1_train })
56  test_df1 = pd.DataFrame({ 'text_cleaned': X1_test, 'target': y1_test
    ↪  })
57
58
59  N_ITER = 10
60
61  gc.collect()
62  torch.cuda.empty_cache()
63
64  for i in range(0,N_ITER):
65      model1 = ClassificationModel(
66      "bert", "bert-base-multilingual-cased",
67          use_cuda = True,
68          args=train_args
69      )
70      model1.train_model(train_df1)
71      result1, model_outputs1, wrong_predictions1 =
        ↪  model1.eval_model(test_df1, f1=f1_multiclass,
        ↪  acc=accuracy_score)
72      print(result1)
73      if(i<limitsave):
74          torch.save(model1, 'model1'+str(i)+'.pt')
75      del model1
76      gc.collect()
77      torch.cuda.empty_cache()
78      if(i==0):
```

```
79          dfResultsModels1 = pd.DataFrame.from_dict(result1,
            ↪  orient="index").T
80      else:
81          dfResultsModels1b = pd.DataFrame.from_dict(result1,
            ↪  orient="index").T
82          dfResultsModels1 = dfResultsModels1.append(dfResultsModels1b)
83
84  dfResultsModels1Trans = pd.DataFrame(dfResultsModels1.mean(axis=0)).T
85  dfResultsModels1Trans['f1'] = calcule_f1(dfResultsModels1Trans)
86
87  dfResultsModelsTotal = dfResultsModels1Trans.copy()
88  dfResultsModelsTotal.to_csv('dfResultsModelsTotal-1.csv')
89
90
91  dfResultsModels1gTrans =
    ↪  pd.DataFrame(dfResultsModels1g.mean(axis=0)).T
92  dfResultsModels1gTrans['f1'] = calcule_f1(dfResultsModels1gTrans)
93  dfResultsModelsTotal =
    ↪  dfResultsModelsTotal.append(dfResultsModels1gTrans)
94  dfResultsModelsTotal.to_csv('dfResultsModelsTotal-1.csv')
95
96  indexBERT =
    ↪  ['BERT','RoBERTa','DistilBERT','CamemBERT','Albert','Flaubert']
97
98  #dfResultsModelsTotal.reindex(indexBERT)
99  dfResultsModelsTotal = dfResultsModelsTotal.reset_index(drop=True)
100 dfResultsModelsTotal.index = indexBERT
101 dfResultsModelsTotal
102
103 dfResultssModelsTotal1 = dfResultsModelsTotal.copy()
104
105 dfResultssModelsTotal1
```

Anexo B

Aportaciones científicas

BERT Model-Based Approach For Detecting Categories of Tweets in the Field of Eating Disorders (ED)

1st José Alberto Benítez-Andrades
SALBIS Research Group
Dept. of Electric, Systems and Automatics Engineering
University of León, Campus of Vegazana s/n
León, Spain
jbena@unileon.es

2nd José Manuel Alija-Pérez
Dept. of Mechanical, Computer Science, and Aerospace Engineering University of León, Campus of Vegazana s/n
León, Spain
jmalip@unileon.es

3rd Isaías García-Rodríguez
SECOMUCI Research Group
Dept. of Electric, Systems and Automatics Engineering
University of León, Campus of Vegazana s/n
León, Spain
isaias.garcia@@unileon.es

4th Carmen Benavides
SALBIS Research Group
Dept. of Electric, Systems and Automatics Engineering
University of León, Campus of Vegazana s/n
León, Spain
carmen.benavides@unileon.es

5th Héctor Alaiz-Moretón
SECOMUCI Research Group
Dept. of Electric, Systems and Automatics Engineering
University of León, Campus of Vegazana s/n
León, Spain
hector.moreton@unileon.es

6th Rafael Pastor Vargas
Communications and Control Systems Department
Spanish National University for Distance Education (UNED)
Madrid, Spain
rpastor@scc.uned.es

7th María Teresa García-Ordás
SECOMUCI Research Group
Dept. of Electric, Systems and Automatics Engineering
University of León, Campus of Vegazana s/n
León, Spain
mgaro@unileon.es

***Abstract*—Eating disorders (ED) are among the most widespread mental illnesses in our society today. This research work presents the study of deep learning models applied to the domain of eating disorders. For this purpose, a collection of messages from the social network Twitter was compiled using web scraping techniques. After collecting a total amount of 1,085,957 tweets, a subset of 2,000 tweets was manually classified. This classification made it possible to differentiate tweets written by people who suffer or have suffered from an ED from those written by people who have not suffered from an ED. After this, 6 predictive models based on Bidirectional Encoder Representations from Transformers (BERT) were created and a comparison was made by evaluating which model scored the best. The best scoring model was RoBERTa using the pre-trained roberta-base model with an accuracy of 87.5%.**

***Index Terms*—NLP, Twitter, eating disorders, deep learning, BERT**

I. Introduction

A large part of today's society places great importance on physical appearance. This fact causes many people to spend a great deal of time and effort on improving their appearance, especially women [1], [2]. Some of the messages sent by magazines and the media often include ideas that induce the purchase of products related to aesthetics and beauty, including clothing (following certain trends), anti-wrinkle creams, and, above all, products that help you lose weight, implying that a slim person will be happier [3]. An example of the strong attachment to this pairing of thinness and happiness are the ongoing debates that have arisen and continue to arise about people who work as models for various advertising campaigns. In contrast, a group of the population has emerged that seems to encourage *curvy* models [4], [5]. For all these reasons, part of the population suffers from mental illnesses related to eating disorders (ED).

ED are life-threatening diseases that are simultaneously psychological and physical in nature [6]. Eating disorders are characterized by a series of abnormal and harmful eating behaviors that are accompanied and motivated by unhealthy beliefs, perceptions and expectations regarding food, weight and body shape [6], [7]. As a general characterization, individuals with ED tend to have difficulty accepting and feeling good about themselves. They tend to think of themselves as "fat" and "ugly" because of their body size and shape, even when this self-criticism is factually inaccurate and untrue. By identifying and defining themselves according to their perceived "fatness," people with eating disorders tend to conclude that they are unacceptable and undesirable and, as a result, feel quite insecure and inadequate, especially in relation to their bodies. For them, controlling their eating behaviors is the logical path in their quest for thinness [6], [7].

Furthermore, despite the complexity of integrating all ED prevalence data, the most recent studies confirm that EDs have a high prevalence worldwide, especially in women. Furthermore, the weighted average point prevalence of ED increased during the study period from 3.5% for the period 2000-2006 to 7.8% for the period 2013-2018. This highlights a real challenge for public health and healthcare providers [8].

For years now, with the emergence of digital social networks such as Twitter, Facebook and Instagram, among others, social media have also been used as a research tool to analyze public debates on a wide variety of topics [9]–[12]. Thanks to these tools and to the emergence of data management and processing strategies using artificial intelligence techniques such as text mining, machine learning or deep learning techniques, these tools can help analyze health care issues such as obesity, eating disorders and derivatives from the patient's perspective [13]–[15]. This research is possible because social networks promote a participatory culture that fosters the creation of communities in which information is shared. This allows them to be used effectively for scientific purposes. In addition, these social media are becoming more and more popular, thus increasing their usefulness as sources for health research [16].

According to different studies, it is possible to assure that social media are currently one of the data sources that can help health research. Health research that makes use of data from social media is continuously growing and is mainly divided into two domains: real-time collection and prediction of diseases and collection of data generated by users and patients on social media. Thanks to social media it is possible to conduct observational studies that can be used to investigate about the information shared by users on a particular topic. Among the main existing social media platforms that are used for health research, the most prominent are: Twitter, Instagram, Youtube, Facebook, Reddit and other interesting blogs and forums [9], [17], [18]. Due to its versatility and ease of data collection and processing, Twitter is the most widely used tool for health-related research [9]. The most researched health-related topics are those dealing with healthcare organization, behavioral medicine, psychiatry, neurology, infectious diseases and oncology [19].

Delving deeper into the studies that have been conducted on data obtained from social media and ED, it is possible to find research related to the detection of people for and against some ED [20]–[22], detection of informative and non-informative tweets [23], [24], and even the detection of communities of people with ED [22], [25]. However, there is still the possibility of creating predictive models using Twitter data that provide interesting information about ED.

In this paper we propose creating four predictive models related to eating disorders using an own dataset obtained from the social network Twitter. To do so, we suggest the use of a BERT (Bidirectional Encoder Representations from Transformers) model-based approach to classify whether a tweet has been written by someone suffering from an eating disorder or not. Different BERT models will be studied and the one which obtains the best results will be determined.

II. Material and Methods

In section A, the dataset on ED tweets is discussed. In section B, the background of BERT (Bidirectional Encoder Representations from Transformers) models is presented. In section C, we discuss how BERT models work and in section D we show the hardware and software setup used for the experiments performed.

In Figure 1 the outline of the work carried out for this research can be seen.

A. Collecting dataset about eating disorders

A tool called T-Hoarder [26] was used to collect the tweets. This tool allows you to select whether you want to obtain tweets related to certain keywords or from a specific user, via streaming. Thanks to its ability to obtain live data, it was possible to obtain tweets that could later be deleted.

The configuration of the collection of tweets was as follows:

- All tweets containing the following keywords were collected: anorexia, anorexic, dietary disorders, inappetence, feeding disorder, food problem, binge eating, anorectic, eating disorders, bulimia, food issues, loss of appetite, food issue, food hater, eat healthier, disturbed eating habits, abnormal eating habits, abnormal eating habit, binge-vomit syndrome, bingeing, bulimarexia, anorexic skinny, eating healthy.
- A subset of the collected tweets were manually labelled by an expert in the field by categorising them into two categories.
  - Tweets written by a person suffering from eating disorders.
  - Tweets written by people without eating disorders

  To perform this categorisation it was necessary to access one by one each user profile that sent such a tweet.

B. Background for BERT-models

The BERT model has caused a sensation in the machine learning community by showing the latest results in a variety of NLP (Natural Language Processing) tasks, including

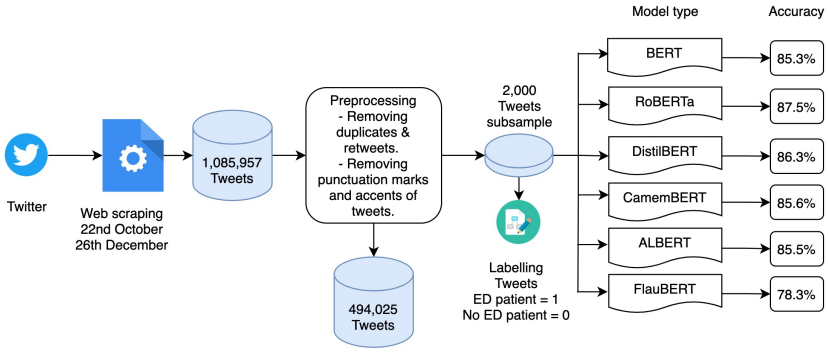

Fig. 1. Outline of the work carried out.

question answering (SQuAD v1.1), natural language inference (MNLI), etc [27].

One of the main technical innovations added in BERT is the application of bidirectional training of Transformer, a popular attention model, to language modelling. Previously, text sequences were examined from left to right or combined left-to-right and right-to-left training. The results obtained by [27] show that language models that are trained bidirectionally have a deeper sense of context and language flow than one-way language models. In addition, the researchers highlight a novel technique called Masked LM (MLM). This technique allows bidirectional training of models that were previously impossible to train with the traditional method.

In the field of computer vision, the importance of transfer learning has been repeatedly demonstrated. This learning occurs after pre-training a neural network on a known task, e.g. ImageNet, and subsequent processing to improve the model. By using a trained neural network as a basis, a new model with a specific purpose is achieved.

Some NLP researchers have shown that this approach can also be used in NLP tasks. By using a pre-trained neural network, it is possible to embed words that are then used as features of NLP models.

C. How BERT models work

BERT models are based on an attentional mechanism that is able to learn contextual relationships between words or sub-words within a text. This mechanism is known as Transformer [28] and has two mechanisms: one capable of reading an input text known as a classifier and another one that makes a prediction for the task known as a decoder. In this case, for BERT models, out of the 2 that Transformer offers, only the encoding mechanism is needed.

The main advantage of Transformer over directional models is that, unlike the latter, the encoder reads the whole sequence of words at once, from left to right and from right to left, not in two different sequences. This is why BERT is considered bidirectional, although it would be more accurate to say that it is non-directional. Thanks to this feature, models can learn the context of a word based on all the words surrounding it, not just those on either side.

When training linguistic models, one of the challenges is to define a clear prediction target. Models that predict the next word in a sequence with a directional approach are inherently limited to context learning. In these cases, BERT makes use of two approaches known as Masked LM (MLM) [29] and Next Sentence Prediction (NSP) [30].

D. Hardware and Software used for the experiments

To perform the pre-processing of the tweets and apply the BERT models, a Jupyter notebook, Python 3.6 was used and run on a computer with the following characteristics: Intel(R) Core(TM) i7-9700K CPU @ 3.60GHZ, 32.0GB RAM and an NVIDIA GeForce RTX 2080 graphics card.

The spaCy and nltk libraries were used to pre-process the text content.

III. Experiments and Results

A. Data collection and pre-processing

The data were collected between October 20, 2020 and December 26, 2020, achieving a total of 1,085,957 tweets.

A preprocessing of the data was carried out that included the following tasks:

- Deletion of all tweets that had the same message.
- Deletion of those retweets for which we did not have the original tweet.
- Deletion of tweets that were not correctly collected.
- Mentions were removed.
- Accents in the text as well as punctuation marks have been removed.
- Stopwords were also removed.

Following this procedure, 494,025 valid tweets were obtained. This dataset is available in an open repository at the

following link: https://www.kaggle.com/jabenitez88/eating-disorders-tweets.

After this pre-processing, a subset of 2,000 tweets was created and tagged in the two categories mentioned above: tweets written by people with ED and tweets written by people without ED. The distribution of categories was 51.5% of tweets written by people who did not have ED versus 48.5% of people who did have ED.

B. Implementation of BERT models

ClassificationModel from the *simpletransformers* library (available at https://github.com/ThilinaRajapakse/simpletransformers) was used to classify the model. This model was used sinceit is suitable for classifying text in binary form within this library.

The scikit-learn library was used to split the dataset using *train_test_split* and also to obtain the f1 score and accuracy score metrics of the generated classification models.

The data were divided into two subsets, 70% training data and 30% validation data.

The following training arguments were assigned:

- reprocess_input_data: True
- fp16: false
- num_train_epochs: 4

Six ClassificationModels were implemented with the above-mentioned arguments and having chosen six different model types by making use of six pre-trained models, which are explained in section III-C.

C. Results

Six neural networks were trained using different types of BERT models (BERT, RoBERTa [31], DistilBERT [32], CamemBERT [33], ALBERT [34] and FlauBERT [35]), all of them using pre-trained models. In Table I it is possible to see the results obtained with each of them.

TABLE I
RESULTS OF DIFFERENT BERT MODELS.

Model type	Pre-trained model	tp	fp	tn	fn	f1	acc
BERT	bert-based-multilingual-cased	245	53	267	35	0.848	85.3%
RoBERTa	roberta-base	247	42	278	33	0.868	87.5%
DistilBERT	distilbert-base-cased	244	46	274	36	0.856	86.3%
CamemBERT	camembert-base	223	29	291	57	0.838	85.6%
ALBERT	albert-base-v1	245	52	268	35	0.849	85.5%
FlauBERT	flaubert_base_cased	203	53	267	77	0.757	78.3%

IV. DISCUSSION AND CONCLUSIONS

By means of this research, a series of messages have been collected through the social network Twitter, all of them related to eating disorders. The initial aim of the research presented in this manuscript was to generate predictive models capable of classifying tweets by determining whether they were written by people with eating disorders or without them.

After preprocessing the data and manually labeling a subset, this was subsequently preprocessed. Once this was done, the data were split into training and test sets and 6 different BERT models were trained. Out of the trained models it was observed that the best scoring model was the one known as RoBERTa using the pre-trained roberta-base model (acc = 87.5%), followed in second place by DistilBERT (acc = 86.3%) and in third place by CamemBERT (acc = 85.3%).

It is observed that, despite having a data sample of 2,000 Tweets, the performance of the different BERT models applied is quite good, which is very promising for the development of future predictive models within the field of ED and other mental illnesses, among others.

Future lines of research include further testing by adjusting more hyperparameters of the models, performing classification tests with other categories and comparing these results with more traditional neural networks in the NLP domain.

ACKNOWLEDGMENT

My special thanks to Mariluz Congosto for her help in realising this project. Without the t-horder, these studies would be much more difficult to carry out.

REFERENCES

[1] D. L. Harris and A. T. Carr, "Prevalence of concern about physical appearance in the general population," *British Journal of Plastic Surgery*, vol. 54, no. 3, pp. 223–226, 2001.

[2] C. Senín-Calderón, J. Gálvez-González, S. Perona-Garcelán, C. Camacho, and J. F. Rodríguez-Testal, "Dysmorphic concern and behavioural impairment related to body image in adolescents," *International Journal of Psychology*, vol. 55, no. 5, pp. 832–841, 2020.

[3] C. Lou and C. H. Tse, "Which model looks most like me? Explicating the impact of body image advertisements on female consumer well-being and consumption behaviour across brand categories," *International Journal of Advertising*, pp. 1–27, sep 2020. [Online]. Available: https://doi.org/10.1080/02650487.2020.1822059

[4] A. Izquierdo, F. Plessow, K. R. Becker, C. J. Mancuso, M. Slattery, H. B. Murray, A. S. Hartmann, M. Misra, E. A. Lawson, K. T. Eddy, and J. J. Thomas, "Implicit attitudes toward dieting and thinness distinguish fat-phobic and non-fat-phobic anorexia nervosa from avoidant/restrictive food intake disorder in adolescents," *International Journal of Eating Disorders*, vol. 52, no. 4, pp. 419–427, 2019.

[5] I. Urdapilleta, S. Lahlou, S. Demarchi, and J. M. Catheline, "Women with obesity are not as curvy as they think: Consequences on their everyday life behavior," *Frontiers in Psychology*, vol. 10, no. AUG, 2019.

[6] T. C. Griffen, E. Naumann, and T. Hildebrandt, "Mirror exposure therapy for body image disturbances and eating disorders: A review," pp. 163–174, nov 2018.

[7] H. W. Hoek, "Review of the worldwide epidemiology of eating disorders," *Current Opinion in Psychiatry*, vol. 29, no. 6, pp. 336–339, nov 2016. [Online]. Available: https://journals.lww.com/00001504-201611000-00004

[8] M. Galmiche, P. Déchelotte, G. Lambert, and M. P. Tavolacci, "Prevalence of eating disorders over the 2000-2018 period: A systematic literature review," pp. 1402–1413, may 2019. [Online]. Available: https://academic.oup.com/ajcn/.

[9] K. A. Timmins, M. A. Green, D. Radley, M. A. Morris, and J. Pearce, "How has big data contributed to obesity research? A review of the literature," pp. 1951–1962, dec 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30022056/

[10] N. Carceller-Maicas, "Youth, health and social networks: Instagram as a research tool for health communication," *Metode*, vol. 0, no. 6, pp. 227–233, apr 2016. [Online]. Available: https://ojs.uv.es/index.php/Metode/article/view/6555

[11] J. Lopez-Castroman, B. Moulahi, J. Azé, S. Bringay, J. Deninotti, S. Guillaume, and E. Baca-Garcia, "Mining social networks to improve suicide prevention: A scoping review," *Journal of Neuroscience Research*, vol. 98, no. 4, pp. 616–625, apr 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jnr.24404

[12] R. Ackland, "Social Network Services as Data Sources and Platforms for e-Researching Social Networks," *Social Science Computer Review*, vol. 27, no. 4, pp. 481–492, nov 2009. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0894439309332291

[13] G. Fiumara., A. Celesti., A. Galletta., L. Carnevale., and M. Villari., "Applying artificial intelligence in healthcare social networks to identity critical issues in patients' posts," in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5: AI4Health,*, INSTICC. SciTePress, 2018, pp. 680–687.

[14] N. Musacchio, A. Giancaterini, G. Guaita, A. Ozzello, M. A. Pellegrini, P. Ponzani, G. T. Russo, R. Zilich, and A. de Micheli, "Artificial intelligence and big data in diabetes care: A position statement of the Italian association of medical diabetologists," p. e16922, jun 2020. [Online]. Available: https://www.jmir.org/2020/6/e16922/

[15] G. R. Bauer and D. J. Lizotte, "Artificial Intelligence, Intersectionality, and the Future of Public Health," *American Journal of Public Health*, vol. 111, no. 1, pp. 98–100, jan 2021. [Online]. Available: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2020.306006

[16] D. Arigo, S. Pagoto, L. Carter-Harris, S. E. Lillie, and C. Nebeker, "Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery," *DIGITAL HEALTH*, vol. 4, p. 205520761877175, jan 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29942634/

[17] M. Singh, D. Bansal, and S. Sofat, "Who is Who on Twitter–Spammer, Fake or Compromised Account? A Tool to Reveal True Identity in Real-Time," *Cybernetics and Systems*, vol. 49, no. 1, pp. 1–25, jan 2018. [Online]. Available: https://doi.org/10.1080/01969722.2017.1412866

[18] M. Singh, A. Singh, D. Bansal, and S. Sofat, "An Analytical Model for Identifying Suspected Users on Twitter," *Cybernetics and Systems*, vol. 50, no. 4, pp. 383–404, may 2019. [Online]. Available: https://doi.org/10.1080/01969722.2019.1588968

[19] L. Sinnenberg, A. M. Buttenheim, K. Padrez, C. Mancheno, L. Ungar, and R. M. Merchant, "Twitter as a tool for health research: A systematic review," pp. e1–e8, jan 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27854532/

[20] S. P. Lewis and A. E. Arbuthnott, "Searching for Thinspiration : The Nature of Internet Searches for Pro-Eating Disorder Websites," *Cyberpsychology, Behavior, and Social Networking*, vol. 15, no. 4, pp. 200–204, apr 2012. [Online]. Available: http://www.liebertpub.com/doi/10.1089/cyber.2011.0453

[21] A. Oksanen, D. Garcia, A. Sirola, M. Näsi, M. Kaakinen, T. Keipi, and P. Räsänen, "Pro-anorexia and anti-pro-anorexia videos on YouTube: Sentiment analysis of user responses," *Journal of Medical Internet Research*, vol. 17, no. 11, p. e256, nov 2015. [Online]. Available: https://www.jmir.org/2015/11/e256/

[22] Y. Fettach and L. Benhiba, "Pro-eating disorders and pro-recovery communities on reddit: Text and network comparative analyses," in *ACM International Conference Proceeding Series*. New York, NY, USA: Association for Computing Machinery, dec 2019, pp. 277–286. [Online]. Available: https://dl.acm.org/doi/10.1145/3366030.3366058

[23] S. Zhou, Y. Zhao, J. Bian, A. F. Haynos, and R. Zhang, "Exploring Eating Disorder Topics on Twitter: Machine Learning Approach," *JMIR Medical Informatics*, vol. 8, no. 10, p. e18273, oct 2020. [Online]. Available: https://medinform.jmir.org/2020/10/e18273/

[24] I. Viguria, M. A. Alvarez-Mon, M. Llavero-Valero, A. A. del Barco, F. Ortuño, and M. Alvarez-Mon, "Eating disorder awareness campaigns: Thematic and quantitative analysis using twitter," *Journal of Medical Internet Research*, vol. 22, no. 7, p. e17626, jul 2020. [Online]. Available: https://www.jmir.org/2020/7/e17626/

[25] T. Wang, M. Brede, A. Ianni, and E. Mentzakis, "Social interactions in online eating disorder communities: A network perspective," *PLOS ONE*, vol. 13, no. 7, p. e0200800, jul 2018. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0200800

[26] M. Congosto, P. Basanta-Val, and L. Sanchez-Fernandez, "T-Hoarder: A framework to process Twitter data streams," *Journal of Network and Computer Applications*, vol. 83, pp. 28–39, apr 2017.

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://www.aclweb.org/anthology/N19-1423

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[29] G. Kushilevitz, S. Markovitch, and Y. Goldberg, "A two-stage masked LM method for term set expansion," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, Pennsylvania: Association for Computational Linguistics, Jul. 2020, pp. 6829–6835. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.610

[30] W. Shi and V. Demberg, "Next sentence prediction helps implicit discourse relation classification within and across domains," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 5790–5796. [Online]. Available: https://www.aclweb.org/anthology/D19-1586

[31] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," 2019.

[32] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," 2020.

[33] L. Martin, B. Muller, P. J. Ortiz Suárez, Y. Dupont, L. Romary, É. de la Clergerie, D. Seddah, and B. Sagot, "CamemBERT: a tasty French language model," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, Jul. 2020, pp. 7203–7219. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.645

[34] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," 2020.

[35] H. Le, L. Vial, J. Frej, V. Segonne, M. Coavoux, B. Lecouteux, A. Allauzen, B. Crabbé, L. Besacier, and D. Schwab, "FlauBERT: Unsupervised language model pre-training for French," in *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 2479–2490. [Online]. Available: https://www.aclweb.org/anthology/2020.lrec-1.302