Estudiando Trastornos de la Conducta Alimentaria en Instagram

TREBALL FI DE GRAU DE

Aleix Alonso Manjón

Directora: Ana Freire

Grau en Enginyeria en Informàtica

Curs 2020-2021

upf. Universitat Pompeu Fabra Barcelona

Escola d'Enginyeria

A mi familia y amigos por apoyarme siempre,
sobre todo en estos últimos años.

Agradecimientos

Me gustaría agradecer a todas las personas que me han apoyado durante la elaboración de este proyecto. En especial, a mi tutora Ana Freire, que sin su guía, consejos, y ayuda, este trabajo sería muy diferente. Gracias Ana por acompañarme durante este trayecto.

También agradecer a la Universidad Pompeu Fabra por los estudios recibidos en los últimos años.

Y por último, a mi familia, amigos y compañeros que aportando sus diferentes granitos de arena me han convertido en la persona que soy hoy.

Resumen

El extensivo uso de las redes sociales provoca que puedan aparecer comunidades peligrosas para determinados colectivos vulnerables. El objetivo de este trabajo es crear un clasificador de perfiles de Instagram que determine si son dañinos para personas que se encuentran en riesgo de sufrir un trastorno de la conducta alimentaria.

Existen diferentes tendencias en Instagram que promueven un estilo de vida saludable, haciendo hincapié en la alimentación, dietas, y deporte, aunque parezcan tendencias positivas, estas no se centran tanto en los beneficios para la salud de estos estilos de vida si no en la apariencia física que puedes conseguir. Este proyecto quiere estudiar si hay personas en riesgo de sufrir un trastorno de la conducta alimentaria que sigan estos perfiles y les pueda aparecer o agravar esta condición.

Abstract

The extensive use of social networks causes dangerous communities for certain groups of people to appear. The goal of this project is to create an Instagram profiles classifier that determines if they are harmful to people who are at risk of suffering an eating disorder.

There exist different Instagram trends that promote a healthy lifestyle, basing it on food, diet, and sport, although they seem positive trends, they do not focus much on the health benefits of these lifestyles but on the physical appearance that you can achieve. This project wants to study if there are people at risk of suffering an eating behavior disorder who follow these profiles and may appear or aggravate this condition.

Resum

L'extensiu ús de les xarxes socials provoca que puguin aparèixer comunitats perilloses per a determinats col·lectius vulnerables. L'objectiu d'aquest treball és crear un classificador de perfils d'Instagram que determini si són nocius per a persones que es troben en risc de sofrir un trastorn de la conducta alimentària.

Existeixen diferents tendències en Instagram que promouen un estil de vida saludable, posant l'accent en l'alimentació, dietes, i esport, encara que semblin tendències positives, aquestes no se centren tant en els beneficis per a la salut d'aquests estils de vida si no en l'aparença física que pots aconseguir. Aquest projecte vol estudiar si hi ha persones en risc de sofrir un trastorn de la conducta alimentària que segueixin a aquests perfils i els hi pugui aparèixer o agreujar aquesta condició.

Tabla de contenidos

1. Motivación y objetivos

1.1 Motivación

Un trastorno de la conducta alimentaria (TCA) es una enfermedad causada por alteraciones en los comportamientos alimenticios de una persona. Los signos de un trastorno de la conducta alimentaria incluyen la obsesión con la comida y por el peso y forma corporal. Estos trastornos no solo afectan a la salud física, sino también a la mental. [1].

En España, entre 1% y 4% de la población sufre algún tipo de TCA. [2] Los TCA son más frecuentes en personas jóvenes y en mujeres. Aproximadamente el 94% de los afectados son mujeres entre la edad de 12 y 36 años. En España, el 21% de estudiantes universitarias y el 15% de estudiantes universitarios se encuentran en riesgo de sufrir un TCA. [3].

Los trastornos de la conducta alimentaria son enfermedades multifactoriales, lo que significa que no hay solamente un motivo que los ocasionan, son diferentes causas que pueden desencadenar en un trastorno de la conducta alimentaria, y este, a su vez, puede ocasionar la muerte.

Según NEDA (*National Eating Disorder Association*), los medios de comunicación contribuyen al riesgo de sufrir un TCA, diversos estudios muestran la relación de la constante exposición al ideal del cuerpo delgado con el malestar corporal entre las mujeres. [4]. Según el estudio "*Eating Disorders and the Role of the Media*" de Wendy Spettigue "los medios de comunicación son un riesgo causal para el desarrollo de un trastorno de la conducta alimentaria", y en particular se destaca en varios estudios el efecto que las redes sociales pueden causar en estos trastornos, debido a la constante exposición al estándar de belleza de un cuerpo delgado. [5].

El uso de las redes sociales está muy extendido, especialmente en la población adolescente, y la proliferación de diferentes movimientos y perfiles que influyen en la alimentación y estilo de vida de las personas, convierte en una urgencia la necesidad de identificar perfiles que pueden llegar a ser perjudiciales, y conseguir un método para poder prevenir posibles consecuencias futuras.

1.2 Objetivos

a) Objetivos del proyecto.

El objetivo de este estudio es identificar perfiles de Instagram que pueden llegar a influir negativamente en usuarios que sean especialmente vulnerables por estar atravesando un trastorno de la conducta alimentaria, potenciando incluso su condición. Los resultados de este proyecto podrían ser muy útiles a la hora de lanzar mecanismos de protección de usuarios vulnerables, como recomendación de usuarios a seguir que no sean perjudiciales para ellos.

b) Objetivos personales.

A nivel personal, este proyecto me ha aportado conocimientos sobre análisis de datos de redes sociales y aprendizaje automático. También he podido aportar mi granito de arena a la caracterización de una enfermedad tan seria y mostrar que hay que tener mucha conciencia de lo que se publica y consume en redes sociales, y que nuestras acciones pueden llegar a ser perjudiciales para determinadas personas.

2. Consideraciones éticas

Teniendo en cuenta las implicaciones éticas de este proyecto, ha sido sometido a evaluación por la Comisión Institucional de Revisión Ética de Proyectos (CIREP-UPF), y está aprobado con el número de referencia 162.

3. Investigaciones previas en detección de TCA

Diversos estudios han analizado el contenido de las redes sociales para encontrar patrones que ayuden a prevenir los trastornos de la conducta alimentaria. Algunos estudios como el paper "*Analysis of Twitter to Identify Topics Related to Eating Disorder Symptoms*" de Sicheng Zhou et al. [6] la intención de este estudio era analizar contenido de redes donde se hable sobre trastornos de la conducta alimentaria y ver si se podía encontrar una forma de prevenirlos, se utilizó la red social Twitter donde fueron evaluados aproximadamente 18.000 *Tweets* relacionados con TCA e identificados 20 temas, en ámbitos relacionados a la salud, la extracción de datos de Twitter es una técnica muy utilizada, este estudio determina que muchos de los comportamientos relacionados con los TCA en internet tienen una similitud a los TCA en la vida real y por lo tanto debería aumentarse la prevención e intervención en estos casos. En el paper de Ling He y Jiebo Luo llamado "*What makes a pro-eating disorder hashtag*" [7] se entrenó un classificador para identificar publicaciones en Tumblr y usuarios en Twitter que promueven y están a favor de los TCA, este estudio se realizó en las redes sociales Tumblr y Twitter porque son dos plataformas donde los *hashtags* relacionados con trastornos de la conducta alimentaria no están prohibidos, también se da hincapié en el preocupante hecho de que en esas redes sociales hay un movimiento en aumento que dice que tener un trastorno de la conducta alimentaria es un estilo de vida en vez de una enfermedad, durante este estudio con datos de Instagram, no se ha sido capaz de encontrar una comunidad con actitudes tan peligrosas, ya que Instagram es una red social muy estricta, pero se ha sido capaz de encontrar publicaciones que pueden inducir al inicio de un trastorno de la conducta alimentaria. En el estudio de Ling He y Jiebo Luo han obtenido un clasificador de precisión del 68% para Tumblr y otro con una precisión del 92% para Twitter y encontraron patrones similares entre ambas redes sociales.

Otros trabajos relacionados con este tema y basados en Instagram son el paper "*Strong beats skinny every time*" de 2016 por Grace Holland BPsych et al. [8] en este estudio, analizaron la tendencia *fitspiration* (*fitness and inspiration*) de Instagram para ver si podía producir trastornos en la alimentación y ejercicio de mujeres. Esta tendencia intenta mediante imágenes y textos inspirar a llevar una

vida saludable con ejercicio y alimentación sana, como alternativa al movimiento *thinspiration* (*thin and inspiration*) que promociona la delgadez, pero ese movimiento, que a priori parece inofensivo, tiene elementos que nos hace pensar que puede ser un movimiento similar al *thinspiration* pero enmascarado en un movimiento sano, ya que la mayoría de publicaciones sólo enseñan un tipo de cuerpo delgado y musculado, y son publicaciones principalmente centradas en la imagen corporal y no tanto en los beneficios de salud que conlleva practicar deporte y tener una alimentación sana. En el estudio, se analizaron 203 mujeres, 101 utilizaban Instagram para subir contenido de *fitspiration*, y 102 utilizaban la plataforma para subir contenido de sus viajes. Se determinó que las mujeres en el primer grupo, tenían más tendencia a querer estar delgada y/o musculada, sufrir bulimia y practicar ejercicio compulsivo. 17,5% de las mujeres del primer grupo estaban en riesgo de ser diagnosticadas con un trastorno de la conducta alimentaria, mientras que del otro grupo sólo un 4,5%, esto evidencia que muchas fotos bajo *hashtags* relacionados con comida y ejercicios pueden tener contenido peligroso que pueda inducir a este tipo de trastornos. A pesar de que el estudio no recoge datos de redes sociales, da una esperanza inicial a este proyecto, ya que sí que encuentra una relación entre el tipo de posts que se van a analizar en este estudio y la posible aparición de un TCA.

El paper "*Instagram use is linked to increased symptoms of orthorexia nervosa*" por Pixie G. Turner & Carmen E. Lefevre [9] determina que el uso de las redes sociales tiene efectos negativos sobre la imagen corporal, depresión, comparación social y trastornos de la conducta alimentaria, el estudio se centra en la ortorexia nerviosa, que es la obsesión por alimentarse sano. El estudio está realizado principalmente en Instagram, y los resultados del estudio determinaron que el mayor uso de Instagram estaba relacionado con una mayor tendencia a sufrir ortorexia nerviosa, siendo la red social en la que se ha encontrado esta relación más alta. El estudio fue realizado mediante una encuesta en línea con una muestra de 680 usuarios que seguían a cuentas de alimentación saludable. El estudio concluye y remarca las implicaciones que tienen las redes sociales en nuestras vidas, y el impacto que pueden tener los *influencers* sobre muchas personas.

En el paper "*Predictors of 'Liking' Three Types of Health and Fitness-Related Content on Social Media: A Cross-Sectional Study*" por Elise R Carrotte et al. [10] se

indica que los adolescentes que consumen contenido relacionado con la forma física en las redes sociales presentan más posibilidades de sufrir un trastorno de la conducta alimentaria. Esto es debido a que son personas en edad de desarrollar patrones de alimentación y las redes sociales están repletas de contenidos que pueden alterar estos patrones. La intención de ese estudio es descubrir qué tipo de población joven está siguiendo a cuentas relacionadas con salud y alimentación. Se dividieron en 3 grupos, cuentas relacionadas con pérdida de peso y motivación para hacer ejercicio (*fitspiration*), cuentas de detox y cuentas de dietas. Se realizó mediante un cuestionario en línea a una muestra de 1000 personas con una media de edad de 21 años. Las conclusiones que se sacaron es que la mayoría de jóvenes que siguen este tipo de cuentas son chicas adolescentes, y se remarca la importancia de controlar este tipo de publicaciones para que enseñen una imagen responsable del ejercicio y de la alimentación.

Vemos que los diferentes estudios donde se han recolectado datos de redes sociales son principalmente de Twitter y Tumblr, esto es debido a la facilidad que ofrecen estas redes para recolectar sus datos, a través de API (*Application Programming Interface*), y la libertad que tienen los usuarios en ellas, ya que por ejemplo, Instagram no permite publicaciones explícitas que promuevan los trastornos alimentarios [11], ni deja recoger sus datos con tanta facilidad. Es por eso que este trabajo supone un reto, y gracias a los estudios que se han realizado donde se ha estudiado el cómo publicaciones que no son explícitamente animando a los trastornos de la conducta alimentaria también pueden producirlos, nos encontramos como este trabajo tiene sentido.

4. Conceptos previos

4.1 *Web scraping*

El *Web Scraping* es una técnica para extraer información de páginas webs imitando los comportamientos de los humanos navegando en la web. [13].

Las técnicas de *Web Scraping* se basan principalmente en la exploración de páginas web para extraer información a través de los elementos de la estructura de HTML (*HyperText Markup Language*). Un elemento es un bloque de código con una etiqueta inicial y final, y contenido en el interior. Estos elementos representan diferentes partes de una web, y pueden diferenciarse con números de identificación o nombre de clases. A continuación, en la Figura 1, podemos ver un trozo de código HTML de Instagram.

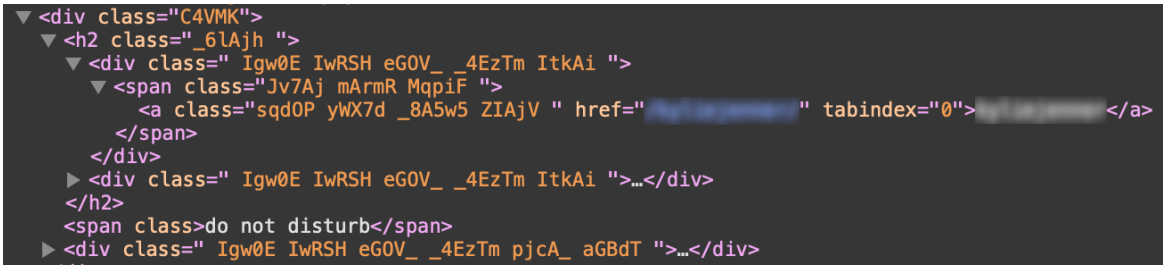

Figura 1: Fragmento de código HTML

En esta imagen vemos un elemento con etiqueta <a>...<a/> que corresponde al nombre de usuario, censurado por motivos de privacidad, y una etiqueta ... con el pie de imagen de la publicación.

Este proyecto se ha realizado con la librería Selenium[1] para Python, que permite acceder de diferentes maneras a estos elementos de HTML. En este caso podríamos acceder al nombre de usuario buscando el elemento con nombre de clase `'C4VMK span'`, especificando así que queremos el primer elemento con etiqueta dentro del elemento que tiene el nombre de clase C4VMK, también podemos acceder especificando el nombre de la clase de ese elemento `'sqdOP.yWX7d._8A5w5.ZIAjV'` o con la dirección XPath, un lenguaje de descripción de rutas utilizando en la web, en este caso sería:

[1] Librería Selenium https://www.selenium.dev

```
/html/body/div[1]/section/main/div/div[1]/article/div[3]/div[1
]/ul/div/li/div/div/div[2]/h2/div[1]/a
```

4.2 Aprendizaje automático

Aprendizaje automático (en inglés, *machine learning*) son una serie de algoritmos de Inteligencia Artificial que permiten a las máquinas aprender a través de datos. Pueden aprender a reconocer patrones para clasificar nuevos conjuntos de datos que no han sido vistos antes por la máquina. [13].

Existen diferentes tipos de aprendizaje automático, los tres principales son aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo, en la Figura 2 podemos ver esta clasificación. El aprendizaje supervisado utiliza un conjunto de datos etiquetados para entrenar un algoritmo, se entrena el algoritmo con un conjunto de entrada y salida etiquetado, para que aprenda a predecir la salida de nuevos datos que le entremos. El aprendizaje no supervisado es la vertiente que utiliza algoritmos de machine learning para analizar conjuntos de datos no etiquetados, este se encarga de descubrir patrones que hay en los datos para entrenar el modelo. [14]. El aprendizaje por refuerzo es similar al aprendizaje supervisado pero este no se entrena con un conjunto de datos iniciales, este aprende utilizando prueba y error. [15].

Para este proyecto, se va a utilizar aprendizaje supervisado, dentro de este, nos encontramos con algoritmos de clasificación y de regresión. [16]. Regresión son algoritmos que para una entrada de datos, el resultado que produce es un número. Clasificación son algoritmos que para una entrada de datos, lo clasifica en una clase, puede ser una clasificación binaria, entre solo dos opciones, o multiclase.

En este proyecto se utilizan diferentes algoritmos de clasificación: Regresión Logística, Árbol de Decisión y Random Forest.

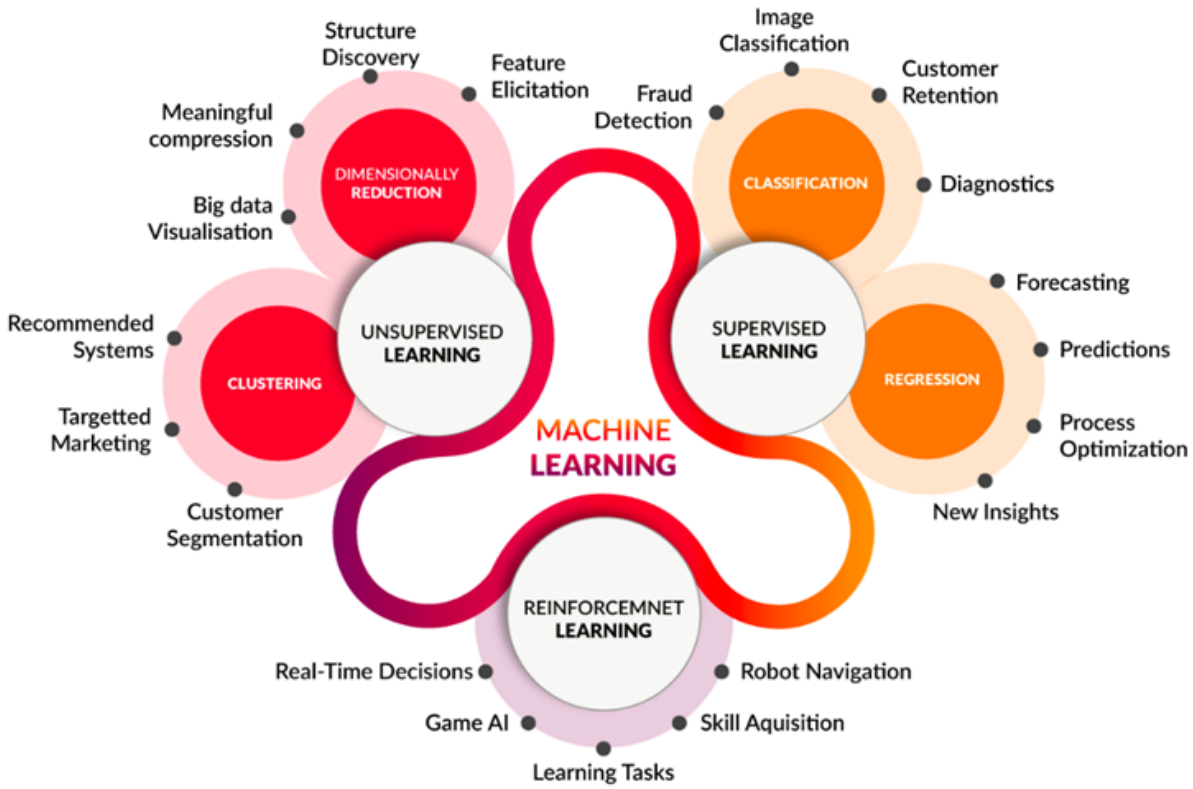

Figura 2: Figura con la clasificación de las subpartes del aprendizaje automático por IBM [16]

El modelo de regresión logística utiliza la función logística para modelar una variable binaria y así poder saber un conjunto de datos a qué clase corresponden

El modelo de Árbol de Decisión utiliza un modelo en árbol para determinar una decisión y las posibles consecuencias. La idea es poder predecir la clase de los datos a través de unas reglas de decisión de los datos con los que se ha entrenado el modelo. [17].

El modelo de *Random Forest* es una extensión del modelo de Árbol de Decisión, este genera un número de árboles de decisión con diferentes subconjuntos de los datos con los que entrenamos para hacer la media y así mejorar la precisión. [18].

Las medidas para evaluar los algoritmos de clasificación son la precisión, sensibilidad (*recall*) y *F1-Score*. La precisión es la capacidad que tiene nuestro clasificador de no clasificar una muestra positiva como negativa. La sensibilidad es la capacidad de encontrar todas las muestras positivas. El *F1-Score* es la media ponderada entre precisión y sensibilidad, el clasificador será mejor cuanto más se acerque a 1. [19].

4.3 *Influencer*

Oxford Languages define a un influencer como alguien que tiene un gran número de seguidores en una red social y expresa opiniones sobre un tema y produce una gran influencia en la gente que lo sigue. Si nos vamos a la herramienta *Google Trends*, podemos observar en la Figura 3 como este término ha ganado importancia recientemente en España, con su resultado de búsquedas más alto en enero de 2021.

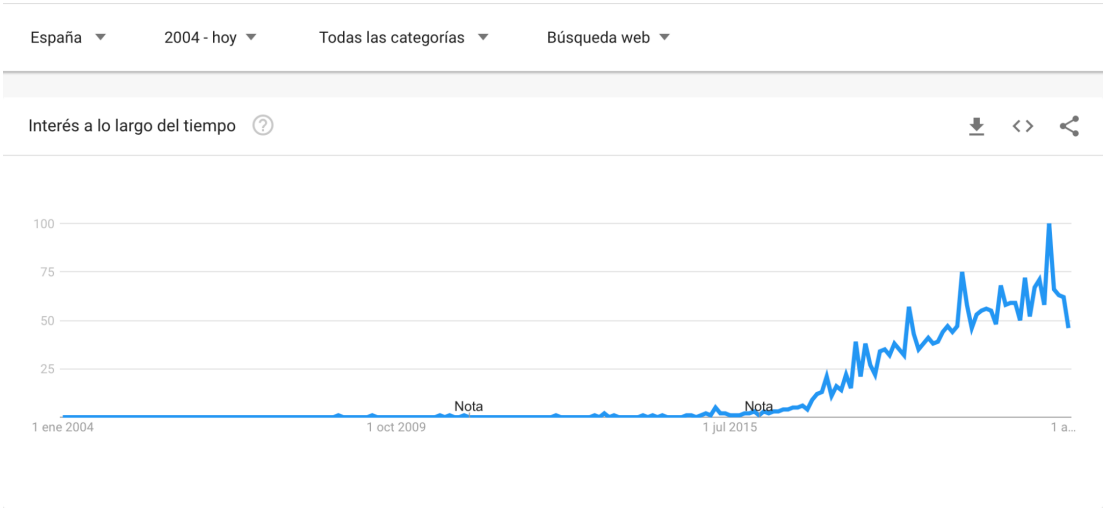

Figura 3: Gráfico de interés de la palabra "*influencer*" según *Google Trends*

Más adelante se propondrá una definición de influencer particularizada para este caso de uso.

5. Propuesta

Recordemos que el objetivo de este proyecto es identificar perfiles de Instagram que pueden llegar a influir negativamente en usuarios que sean especialmente vulnerables por estar atravesando un trastorno de la conducta alimentaria, potenciando incluso su condición.

Para llevar a cabo esta propuesta, se han llevado a cabo las tareas detalladas en las siguientes secciones:

5.1 Creación de un *web scraper* de instagram

La herramienta para realizar el scrapping ha sido desarrollada con el lenguaje de programación Python y con la ayuda de la librería Selenium, una librería que permite crear *scripts* para automatizar la navegación en páginas web. Este es uno de los elementos principales del proyecto, ya que Instagram, a diferencia de otras redes sociales como Twitter o Tumblr, no ofrece una API (*Application Programming Interface*) pública y se ha tenido que recurrir a técnicas de *scraping* para lograr la recolección de los datos.

El *scraper* funciona de la siguiente manera:

1. Introducimos el *hashtag* del que queremos recopilar publicaciones, el *scraper* abre la página de “explorar” de Instagram, y se obtienen las últimas N publicaciones bajo este *hashtag*, incluyendo las 9 publicaciones destacadas: publicaciones populares para un determinado *hashtag*. En la Figura 4 podemos observar un ejemplo de en qué consiste la página “explorar” relativa al *hashtag fitspain*.

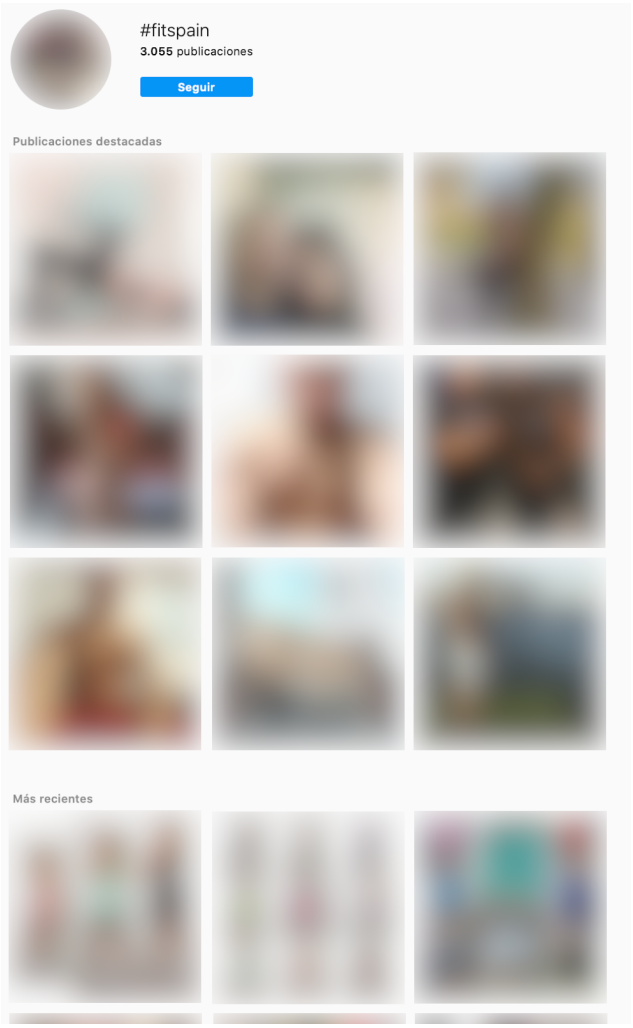

Figura 4: Estructura de la página “explorar” de Instagram

2. Después de obtener los enlaces de las imágenes deseadas, se accede a cada publicación.

3. Para cada publicación se guarda el pie de imagen, el número de likes, el nombre del perfil, que se encripta en MD5 para que no sea posible desencriptar, pero más adelante esto nos ayudará a agrupar las imágenes que sean de un mismo perfil.

4. A continuación, se entra en el perfil que ha publicado esta imagen, se guarda el número total de posts, de seguidores y de seguidos, y guarda el pie de foto de las últimas 25 imágenes.

5. Para finalizar, se accede a la lista de los seguidores y se obtiene la descripción de los últimos N seguidores, en el caso de este proyecto, los últimos 50 seguidores.

A continuación podemos ver una representación visual de todos los elementos que se extraen de una publicación:

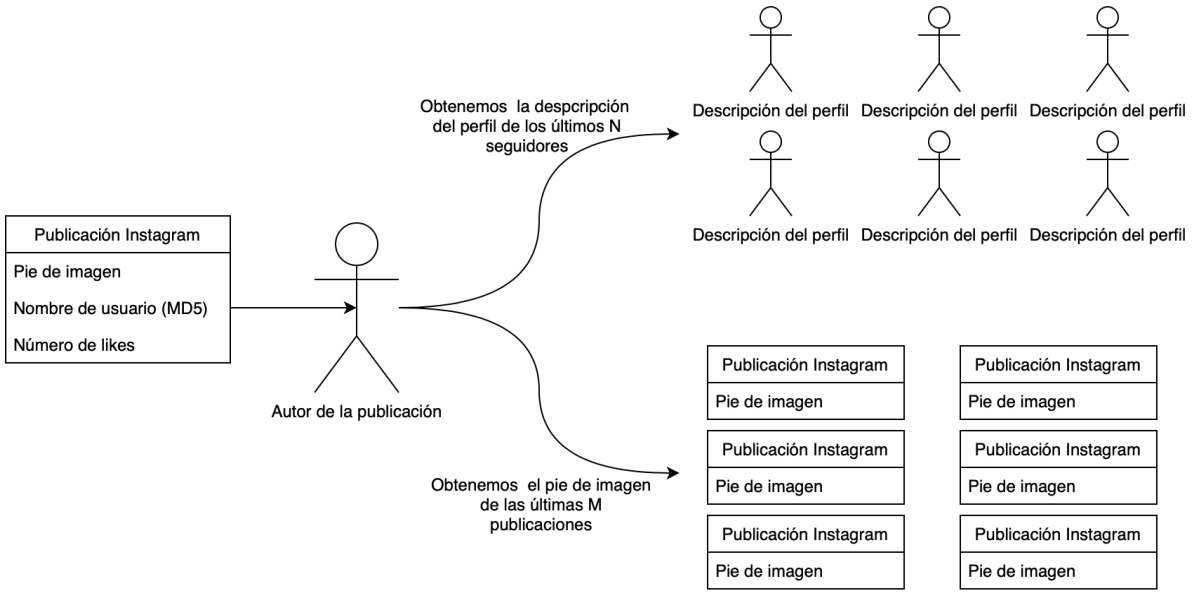

Figura 5: Representación visual de los elementos extraídos de una publicación

Para este proyecto, se han escogido 23 hashtags relevantes relacionados con pérdida de peso, ejercicio extremo o conocidas dietas para adelgazar y veganas (y, a su vez, relacionados con usuarios que padecen un TCA) y se han extraído entre 100-115 fotos de cada uno. Los hashtags escogidos son los siguientes *fitspain, fittrack, realfooding, ketodiet, ketojourney, ketoinspiration, ketolifestlye, vidasaludable, comidareal, perdidadepeso, veganfoodporn, españafit, healthylifestyle, adelgazarcomiendo, herbalife, herbalifenutrition, womensbest, bodytransformation, cleaneating, paleodiet, lowcarb, instanutricion, consejonutricion, habitosaludables, reto1mesconcomidareal.*

Por temas éticos y legales, se ha anonimizado la colección de datos recopilada. Todos los textos han sido anonimizados con una función propia, ya que al principio se utilizó la librería Scrubadub[2] pero realizando diferentes pruebas, se vio que eliminaba palabras que podrían resultar interesantes para este estudio, como las tallas de ropa, es por eso, que con la ayuda de una base de datos de 100.000[3] nombres propios, se anonimizaron los textos. Como se necesitaban que los datos fueran totalmente anónimos, también se anonimizaron menciones a otros usuarios o

[2] Scrubadub: https://pypi.org/project/scrubadub/
[3] Name Dataset de Philippe Rémy: https://github.com/philipperemy/name-dataset

páginas web que podrían estar enlazadas en el post, para así no poder encontrar ninguna relación con la publicación.

El principal problema para la obtención de los datos fueron las dificultades que Instagram presenta a la hora de recolectar datos, es por esto que no se pueden obtener grandes volúmenes de datos, a diferencia de en otras redes sociales como Twitter. Entonces para un proceso que de antemano parece sencillo, se convierte en un trabajo de meses para no ser bloqueado por Instagram. Esto significa que no se puede realizar esta tarea por fuerza bruta, porque la dirección IP (*Internet Protocol*) y la cuenta de Instagram se verían bloqueadas. Así que ya no es un proceso automático, sino uno asistido.

Para no alcanzar el límite máximo de peticiones de IP sobre Instagram, se tenía que ir renovando la IP entre un número determinado de iteraciones y ser cauteloso de no hacer demasiadas peticiones en un periodo muy corto de tiempo. Para lograr esto, entre diferentes acciones se han establecido temporizadores con un tiempo aleatorio, para poder evitar que instagram detecte comportamientos automatizados.

Otra técnica que utiliza Instagram para prevenir el *scraping*, es cambiar el nombre de las clases y de la estructura de la web, así que el código tenía que ir evolucionando y adaptándose a los nuevos cambios.

Debido a estas restricciones, de la colección inicial de *hashtags* solo se pudieron obtener las descripciones de los seguidores de los 12 primeros *hashtags*, y en algunos de estos solo se obtuvo de perfiles influencer, esto es debido a que al principio se empezó con la Idea de obtenerlos todos, pero viendo las dificultades de avanzar, se decidió obtener resultados de perfiles de solo influencer para poder tener representación de todos los *hashtags* y no solo tener un gran volumen de usuarios concentrados en un *hashtag*. Además, se decidió extraer más información de los usuarios influencers ya que es de especial importancia saber si en este tipo de perfiles están sucediendo actitudes que puedan inducir o potenciar un TCA, ya que son perfiles con gran repercusión y un gran alcance.

Para entender esto, hay que entender que entrar en los 50 últimos seguidores de 555 perfiles supone entrar en unos 27.750 perfiles de usuarios. De todos estos, 17.271 tienen descripción, y esto es debido a que existe la posibilidad de que la

cuenta no tenga biografía, es por eso que el tamaño total de biografías es menor al de perfiles visitados.

5.2 Recopilación de un conjunto de datos para la predicción de riesgo de influencia negativa sobre usuarios vulnerables

Tras la recolección de datos, la colección obtenida se formaba de 2872 publicaciones de Instagram, pero no todas han podido ser utilizadas, ya que no se pudo obtener la información de la descripción de los últimos 50 seguidores para todos los usuarios. Se obtuvo la información completa de 793 publicaciones, y después de agruparlas por seguidores únicos, se obtuvieron 555 perfiles únicos, con el pie de imagen de 25 fotos y aproximadamente 50 descripciones de sus últimos seguidores.

Los otros datos no se pudieron utilizar ya que no se disponía de la información de los seguidores, y este dato es crucial para determinar si un perfil puede resultar dañino o no, desafortunadamente, se intentaron diversos procedimientos para extraer el mayor número de datos posibles pero Instagram no lo hizo una tarea sencilla.

a) Elaboración de la *ground truth*

Con el objetivo de predecir si un usuario puede tener un alto riesgo de influencia negativa para posibles seguidores vulnerables, se plantea un problema de clasificación basado en aprendizaje automático supervisado. Las clases que queremos predecir son "*influencer* negativo", si puede llegar a influir negativamente a personas vulnerables que estén atravesando un TCA, o "*influencer* neutro", refiriéndose a usuarios que no suponen una amenaza para sus seguidores.

Para etiquetar el conjunto de datos consideramos que un usuario es dañino si tiene al menos un seguidor que su biografía pertenece a "alto riesgo de TCA" o "riesgo medio de TCA". Para determinar este riesgo en sus seguidores, se escogió un set de palabras clave cuyo uso definiría el riesgo de TCA.

En la categoría de "alto riesgo de TCA" se encuentran palabras como: *SW, CW, GW, thinspo, thynspo, thinspiration, tight, gap, bonespo, eatingdisorder, proana,*

thighgap, mia, ednos, anorexia, ana, fitspo, bulimia, anorexic, tw, anorexianervosa, eupd, tca, ed. Se propuso que SW (*start weight*), CW (*current weight*), and GW (*goal weight*) pertenecían a la categoría de alto riesgo porque podemos observar claramente que la gente que pone eso en su descripción están sobre-preocupadas por su peso e imagen corporal.

En la categoría de "riesgo medio de TCA", encontramos principalmente palabras relacionadas con dieta, ejercicio y pérdida de peso, estas son: *weightloss, kg, lb, lbs, diet, weight, lose, loss, lost, losing, fat, burner, keto, lchf (Low-Carb, High-Fat), omad (One Meal A Day), pounds, ketosis, slim, fit, fitness, calories, ketogenic*.

Como se comentó anteriormente, etiquetamos un perfil como "*influencer* negativo" si tiene al menos un seguidor de su biografía que pertenece a uno de los dos grupos de riesgo de TCA. Si es un *influencer* negativo, el tag *harmful* se establecerá a 1 y si es un *influencer* neutro, es decir, un perfil no peligroso, a 0.

b) Extracción de características

Como era de esperar cuando se recolectaban los datos, diferentes publicaciones corresponden al mismo propietario, y como se quería determinar si un perfil era peligroso o no, se han tenido que agrupar según su nombre de usuario, en este caso a través de la encriptación MD5 (el mismo nombre de usuario producirá la misma encriptación MD5, pero sin poder desencriptar de MD5 al nombre de usuario real). Después de agrupar las publicaciones por nombre de usuario, se eliminaron los pies de imágenes repetidos de cada perfil, para tener un conjunto de pies de imágenes únicos. A partir de aquí, se extrae el conjunto de características que conformarán la entrada a los algoritmos de aprendizaje automático.

Para analizar cada pie de imagen, se utilizaron dos librerías de Python, TextBlob[4], que retorna la polaridad y subjetividad del texto, y NRCLex[5], que retorna las emociones principales del texto, esta librería reconoce las siguientes emociones: *fear, anger, anticipation, trust, surprise, positive, negative, sadness, disgust, joy*. También, para cada perfil se extrajeron los *hashtags* más comunes que utilizaban, como hay tantos *hashtags*, se propusieron diferentes categorías y se utilizaron

[4] https://textblob.readthedocs.io/en/dev/
[5] https://pypi.org/project/NRCLex/

expresiones regulares para poder abarcar diferentes formas de estas palabras claves, las categorías son las siguientes (nótese que se mezclan palabras en castellano con palabras en inglés, porque estas últimas están muy extendidas en redes sociales en su forma anglosajona y aparecen en publicaciones escritas mayormente en castellano):

- Grupo Fit: *fit, sport, entrenamiento, deporte, gym, culturismo, bodybuild, strength, strong, fuerza.*
- Grupo Vegano: *vegan, vegetarian, veggie, vegetal, natural, bio, organic*
- Grupo Dieta: *diet, dieta, keto, lowcarb, lchf, realfood, omad*
- Grupo Salud: *health, sano, sana, saludable*
- Grupo Peso: *weight, peso, calori, fat, kg, kg, pound*.

La estructura del conjunto de datos final es:

- username → El nombre del perfil del usuario, encriptado en MD5
- influencer → Si el perfil es *influencer*, calculado a través del número de "me gustas" de la publicación, el total de publicaciones, número de personas que le siguen y a las que sigue.
- harmful → Si es un "influencer negativo"
- polarity → Si el texto expresa una opinión positiva o negativa
- subjectivity → Si el texto expresa opiniones o sentimientos de quién lo ha escrito
- negative → Si el texto expresa sentimientos negativos
- positive → Si el texto expresa sentimientos positivos
- anticipation → Si el texto expresa sentimientos de anticipación
- joy → Si el texto expresa sentimientos de alegría
- trust → Si el texto expresa confianza
- anger → Si el texto expresa sentimientos de enfado
- fear → Si el texto expresa sentimientos de miedo
- disgust → Si el texto expresa sentimientos de asco
- sadness → Si el texto expresa sentimientos de tristeza
- surprise → Si el texto expresa sentimientos de sorpresa
- fit_group → Si el perfil publica imágenes bajo *hashtags* del "grupo fit"

- vegan_group → Si el perfil publica imágenes bajo *hashtags* del “grupo vegano”
- diet_group → Si el perfil publica imágenes bajo *hashtags* del “grupo dieta”
- health_group → Si el perfil publica imágenes bajo *hashtags* del “grupo salud”
- weight_group → Si el perfil publica imágenes bajo *hashtags* del “grupo peso”

5.3 Estimación del grado de influencia de un usuario de Instagram.

Se ha adaptado la fórmula del paper “*Forecasting Apple Inc Volatility Shares using Twitter Sentiment Analysis*” de Víctor Perez Cester [20] para definir un *influencer*:

$$Engagement_{user} = \frac{(user_{tweets} + user_{likes})user_{followees}}{user^2_{followers}}$$

Como esta fórmula fue creada para twitter, se ha adaptado con los correspondientes campos en Instagram.

$$Engagement = \frac{(user_{posts} + post_{likes})user_{followees}}{user^2_{followers}}$$

Engagement es un término que se utiliza en redes sociales para definir el alcance de los perfiles. Realizando diversas pruebas, se ha podido observar que la región de *influencers* se encuentra en los valores menores a 0.2. En la Tabla 1 podemos ver diversos ejemplos:

Nombre de usuario	Publicaciones	Me gustas	Siguiendo	Seguidores	*Engagement*
no_influyente_1	2753	213	7047	2896	2,50
no_influyente_2	682	121	835	861	0,90
no_influyente_3	144	67	329	332	0,63
no_influyente_4	209	189	1746	1404	0,35
no_influyente_5	15	57	152	184	0,32
influyente_1	8349	8371	2147	431515	0,002
influyente_2	458	4811	322	356191	0,0000134

influyente_3	8478	60657	931	2898534	0,000008
influyente_4	2076	43892	7	1531387	0,0000001
influyente_5	6725	5608814	69	233458475	0,000000007

Tabla 1: Cálculos del *engagement* de diferentes usuarios

Hay que tener en cuenta que en algunos casos puede haber un pequeño error de clasificación, y esto es debido a que como Instagram no libera los datos de los usuarios, es muy difícil obtener la media de “me gusta” de todas las publicaciones, es por eso que en la ecuación como número de “me gusta” se utiliza el de la publicación que se ha obtenido, y puede haber un poco de sesgo si la publicación acaba de ser subida y por lo tanto no tiene los “me gusta” que suelen tener, o si resulta ser una publicación que ha recibido más “me gusta” de lo normal.

5.4 Análisis preliminar de datos

Antes de desarrollar los modelos de predicción del grado de influencia negativa mediante algoritmos de Aprendizaje Automático, se ha realizado un análisis preliminar de los datos.

Del conjunto total de datos, se observan más usuarios influencer que no influencer, esto no nos preocupa ya que este proyecto quería centrarse sobre todo en cuentas que tengan cierta relevancia y lleguen a un número más grande de usuarios.

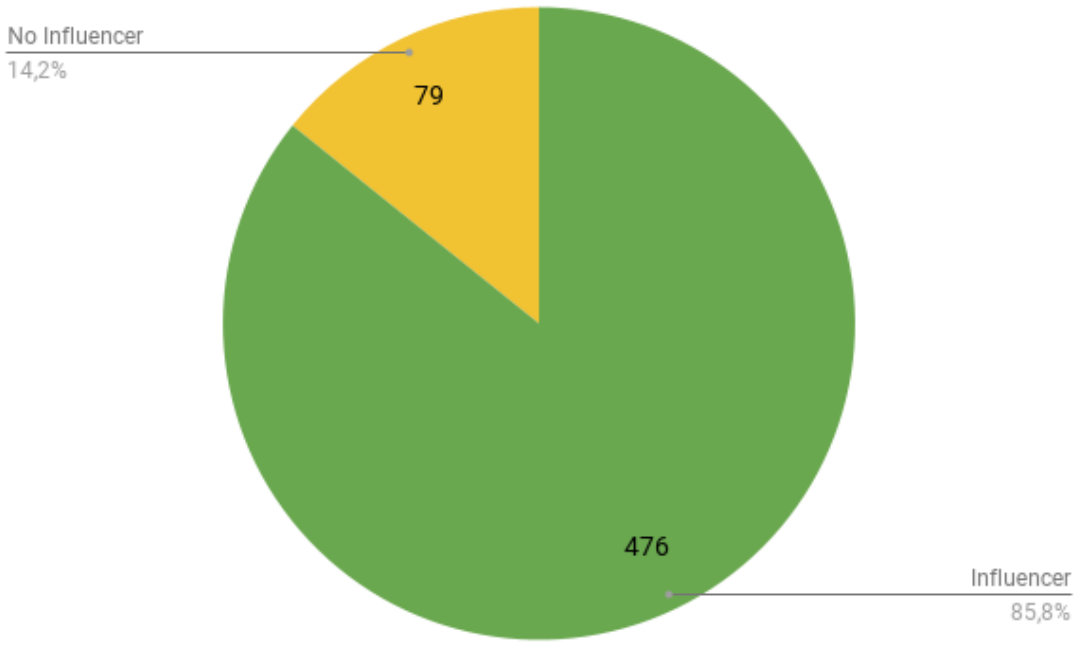

Figura 6: Clasificación en perfiles *influencer* y no *influencer* del total de usuarios obtenidos

De los perfiles totales, se puede observar la separación entre perfiles dañinos y no dañinos.

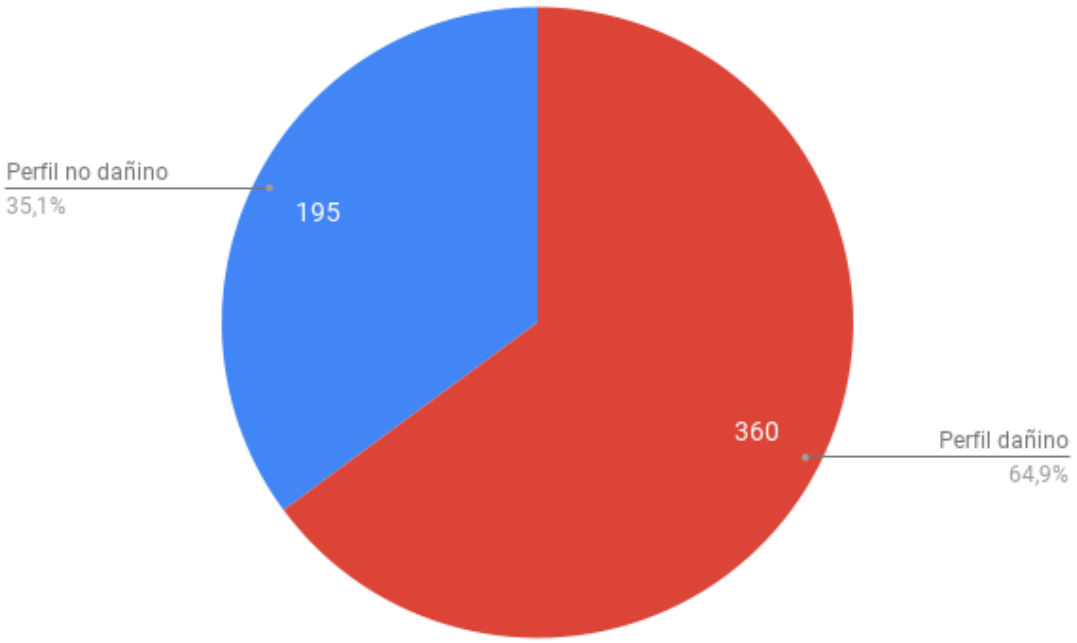

Figura 7: Clasificación en perfiles dañinos y no dañinos del total de usuarios obtenidos

Si se separan los usuarios por *influencer* y no *influencer*, aunque de perfiles no *influencers* se disponen de menos muestras, se observa como la tendencia se mantiene similar.

De los usuario *influencer*, se obtiene las siguiente clasificación:

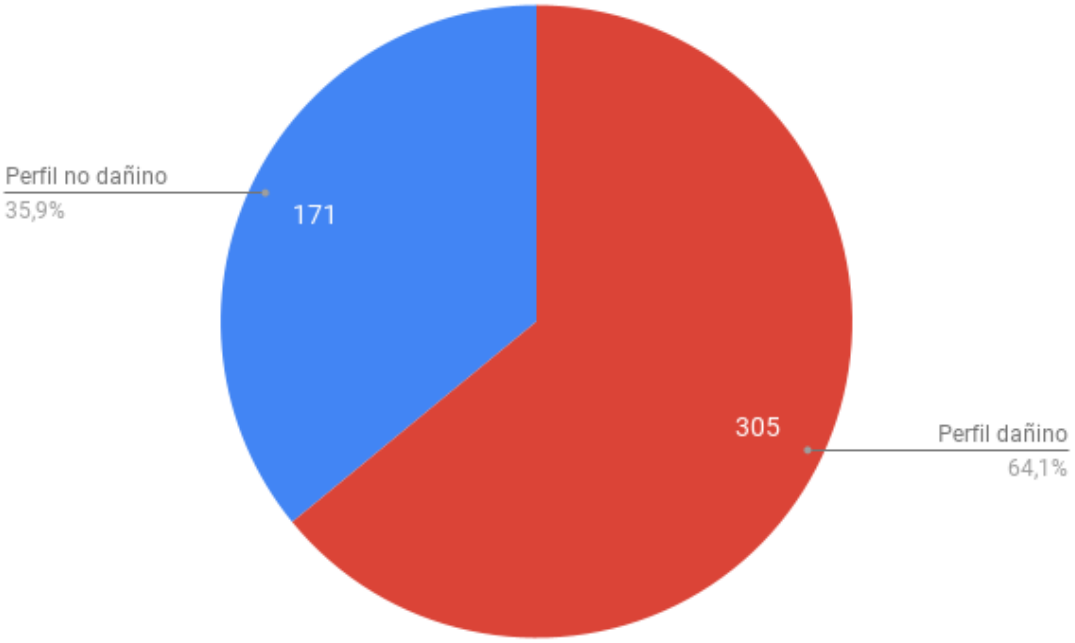

Figura 8: Clasificación en perfiles dañinos y no dañinos de usuarios *influencer*

Y de los perfiles no *influencer*:

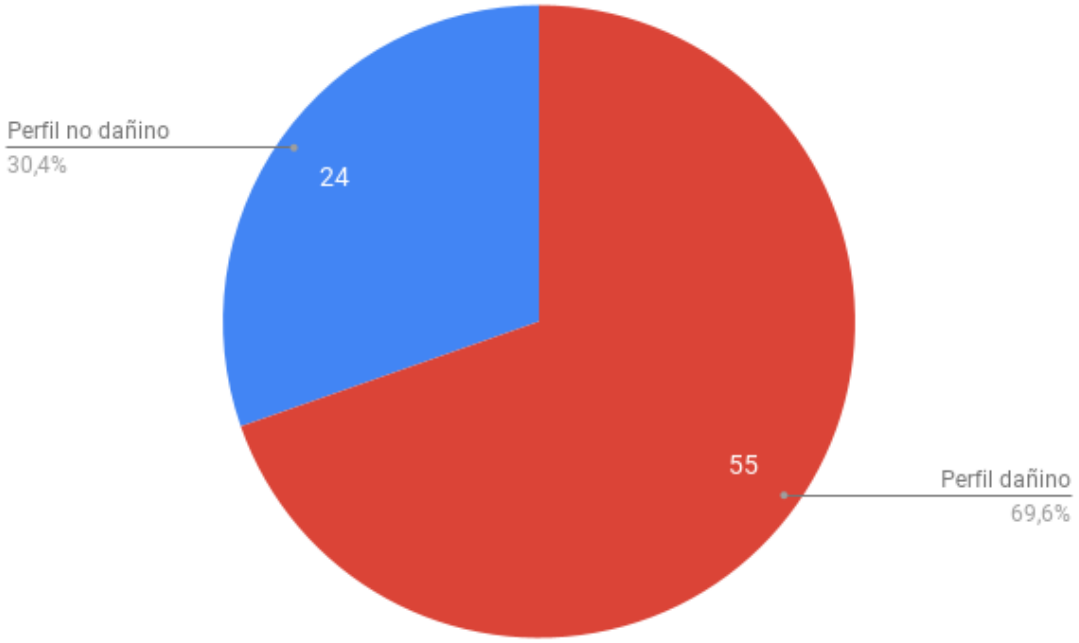

Figura 9: Clasificación en perfiles dañinos y no dañinos de usuarios no *influencer*

Con la ayuda de diversos gráficos, se puede ver como entre los datos obtenidos, se encuentran perfiles que podríamos categorizar como "influencers negativos", como podemos ver en los siguientes gráficos, correspondientes a dos de estos usuarios.

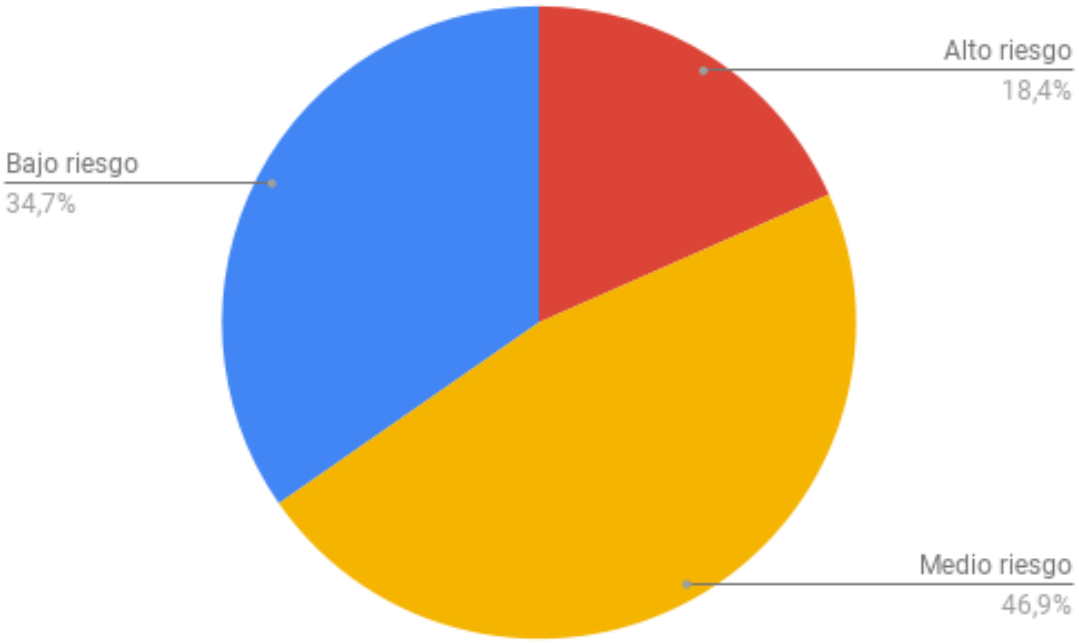

Figura 10: Porcentajes de seguidores por nivel de riesgo del perfil con mayor número de seguidores vulnerables a sufrir un TCA

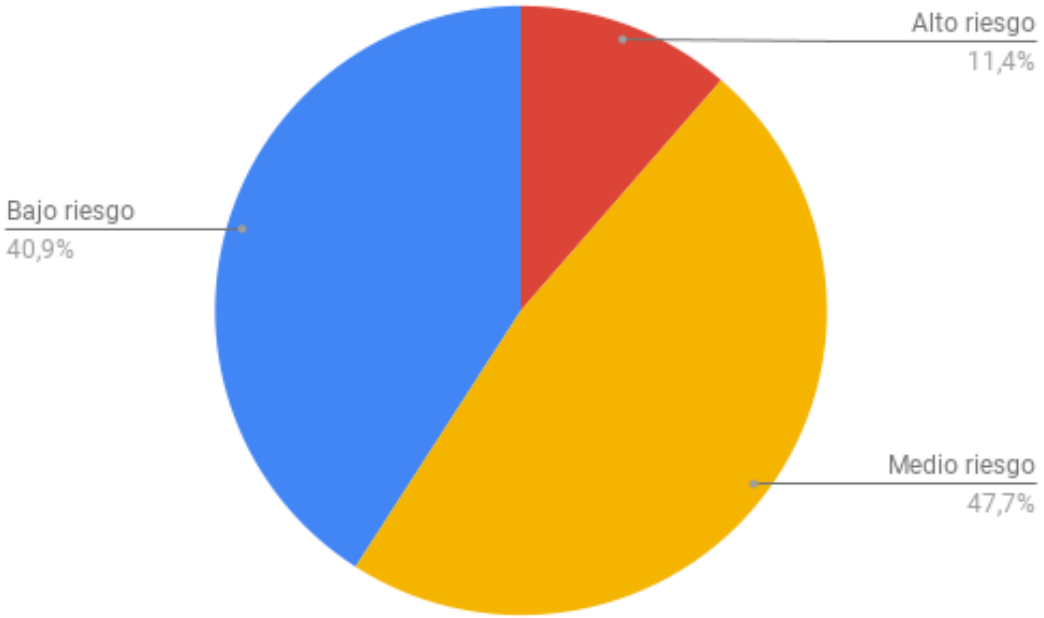

Figura 11: Porcentajes de seguidores por nivel de riesgo del segundo perfil con mayor número de seguidores vulnerables a sufrir un TCA

Estos dos perfiles son los que tienen un porcentaje más alto de seguidores con riesgo alto a tener un trastorno de la conducta alimentaria, también, se observa otro ejemplo de perfil que no tiene seguidores con riesgo alto pero sí que tiene más seguidores con riesgo que seguidores sin riesgo.

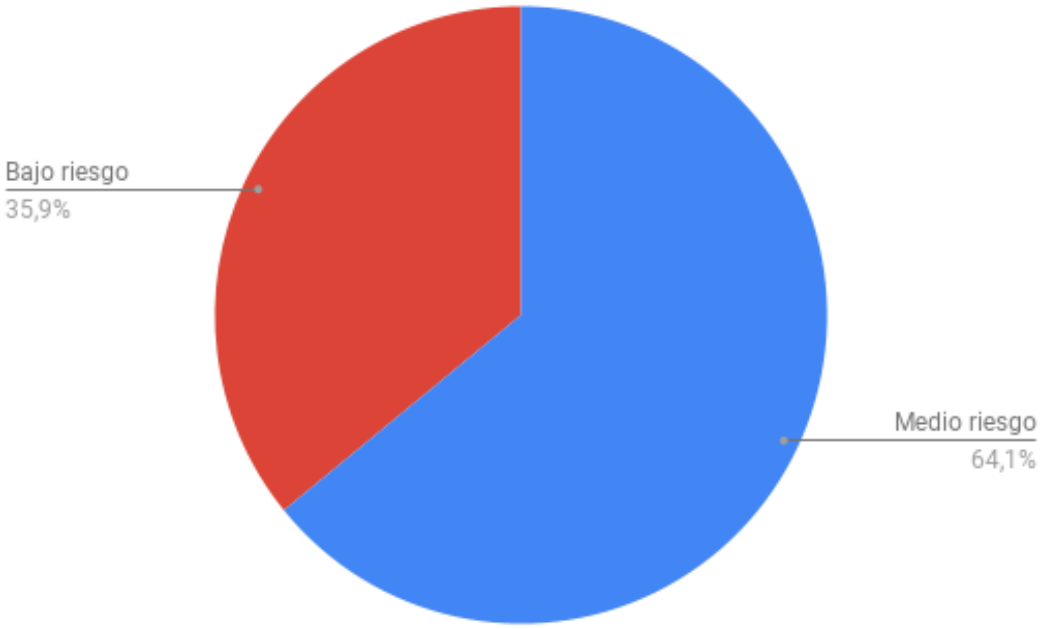

Figura 12: Porcentajes de seguidores por nivel de riesgo de un perfil con más seguidores en riesgo medio que en riesgo alto

Si se calcula la media de todos los perfiles que pueden ser dañinos, se obtienen los siguientes resultados:

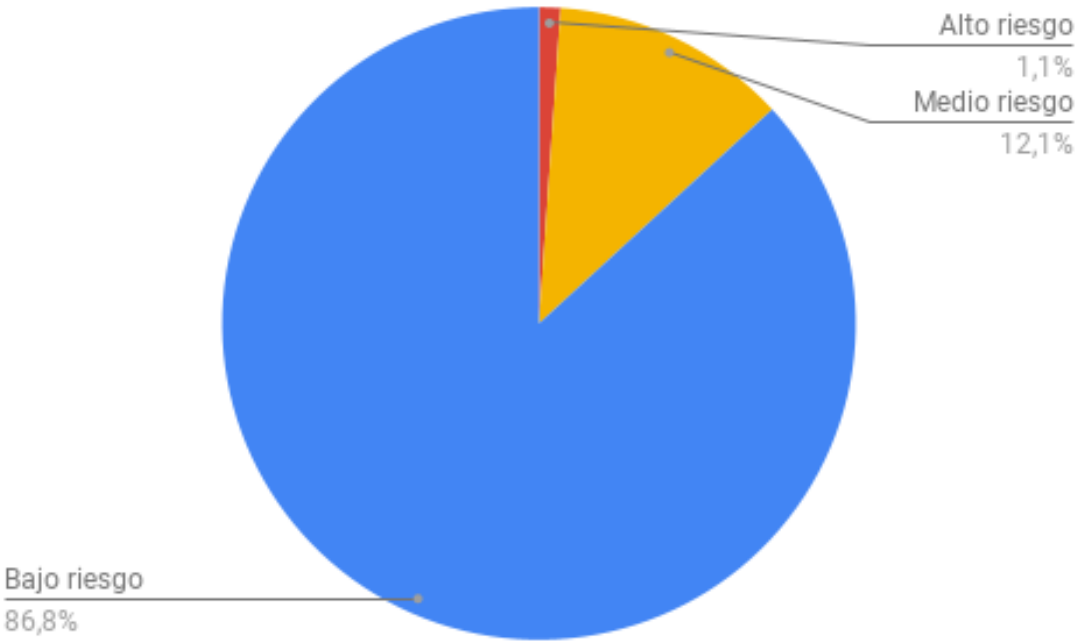

Figura 13: Porcentajes de la media de seguidores por nivel de riesgo de los usuarios que siguen a perfiles peligrosos

Esto significa, que de los 360 perfiles obtenidos que presentan seguidores con un riesgo de sufrir algún tipo de TCA, al menos un 1% presenta riesgo alto, y un 12% riesgo medio, con esto, vemos que de una muestra de 100 seguidores, casi 15 personas están en riesgo de sufrir un TCA. Esto es alarmante si se considera que la muestra de perfiles obtenidos no ha sido muy grande, debido a todas las limitaciones de Instagram, ni se han podido obtener más seguidores por perfil, y estos han sido únicamente los 50 últimos seguidores de un perfil.

En el siguiente gráfico, se observa la media de cada característica del análisis de los textos para “influencers negativos” (*harmful*) e “influencers neutros” (*not harmful*).

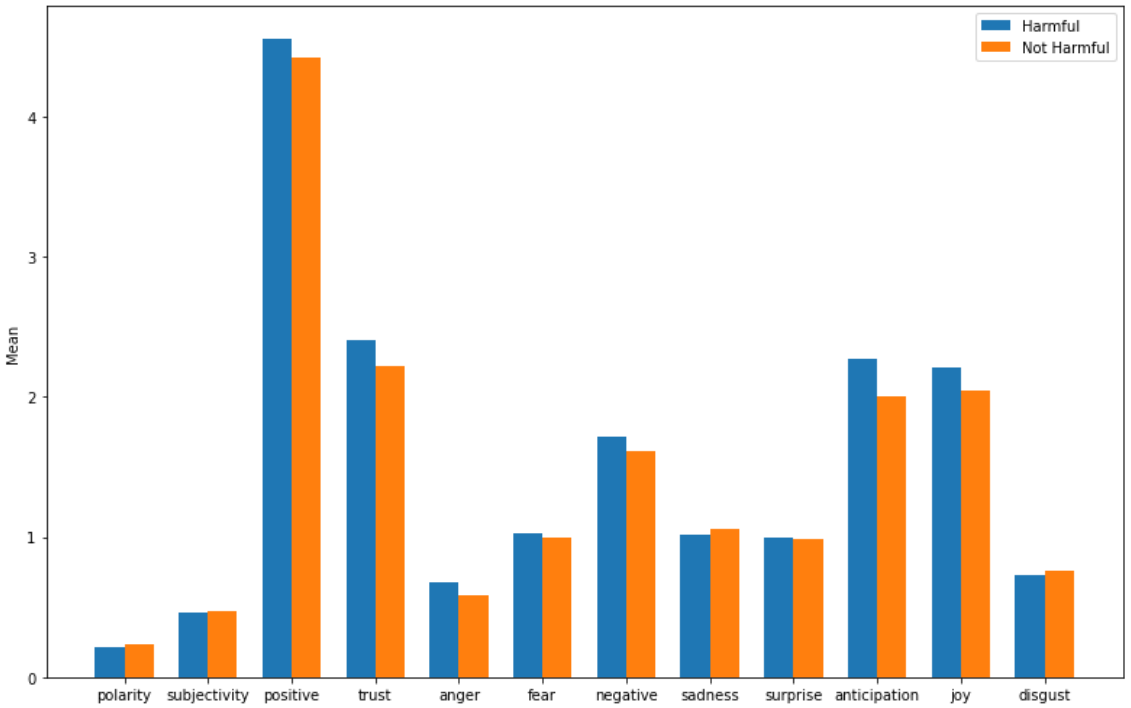

Figura 14: Gráfico comparativo del análisis de emociones de los textos

Y en el siguiente gráfico, se puede ver la media por perfil dañino y no dañino de los diferentes grupos de *hashtags*, sobre todo destaca como los perfiles dañinos suben más fotos bajo los *hashtags* relacionados con deporte, dieta y peso.

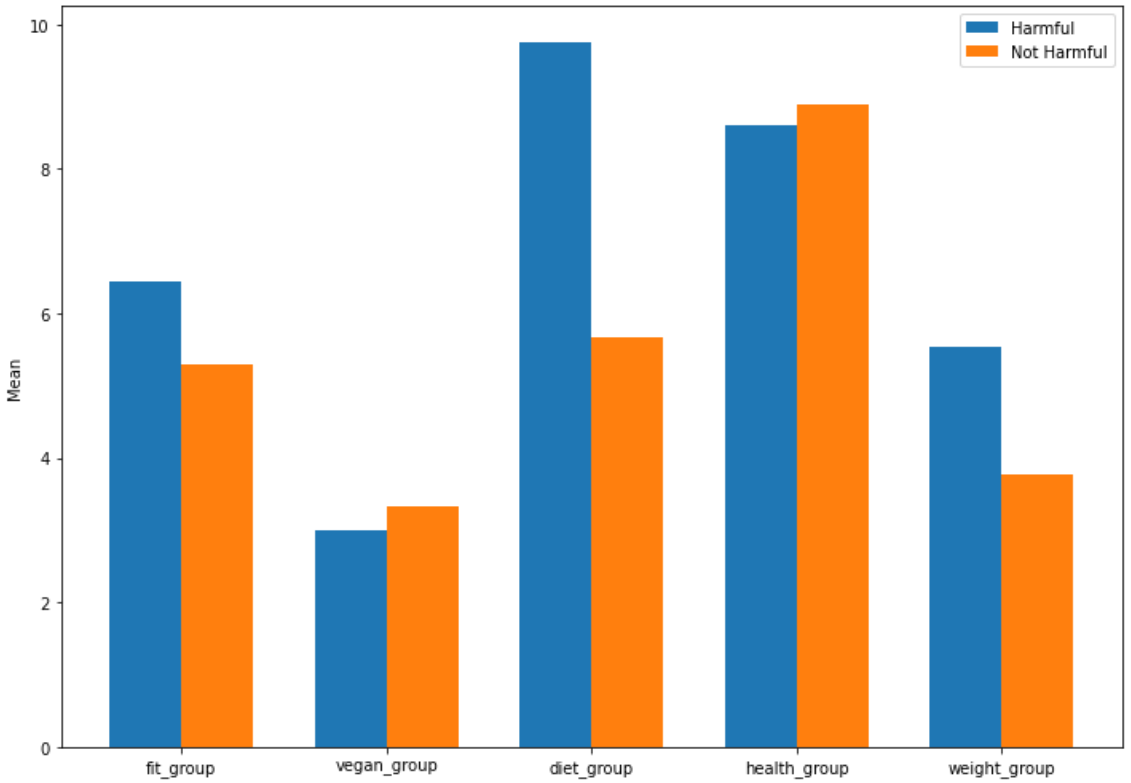

Figura 15: Gráfico comparativo del uso de *hashtags* para los diferentes tipos de perfiles

5.5 Predicción del riesgo de influencia negativa sobre usuarios vulnerables

Una vez entendido los datos que se obtuvieron y procesados para poder analizarlos con Aprendizaje Automático, se utilizaron los métodos de Regresión Logística, Árboles de decisión y Random Forest para establecer una comparación entre ellos. Se observó que los datos obtenidos no estaban balanceados, ya que de 555 perfiles únicos, 360 perfiles eran dañinos y 195 no, si se realizan los algoritmos de aprendizaje automático con un conjunto de datos no balanceados, es muy probable que, a pesar de que los resultados de precisión obtenidos puedan ser altos, lo que esté sucediendo es que siempre se esté asignando las muestras a la clase predominante, ya que al tener pocas muestras de una clase, el algoritmo no aprendería de forma correcta cuál es el límite de decisión para determinar si una nueva muestra es de una clase u otra, por lo tanto, el algoritmo no sabría distinguir entre las dos clases. Existen diferentes alternativas para balancear los datos, siendo una la de duplicar los mismos, esto balancearía los datos pero no porporcionaría información nueva para el modelo, en este caso, para solucionar este problema, se utilizó la técnica SMOTE (*Synthetic Minority Oversampling Technique*), que a partir de los datos de la clase con menor información, en este caso, la de perfiles no dañinos, crea nuevos valores para poder trabajar con un conjunto balanceado. Estos nuevos datos se crean seleccionando muestras de la clase minoritaria del modelo de manera aleatoria, después, busca *k* diferentes muestras cercanas en el espacio, y escoge una de manera aleatoria, y se crea una nueva muestra en un punto aleatorio entre las dos muestras seleccionadas. Este es uno de los métodos más efectivos porque las nuevas muestras obtenidas son muestras cercanas en el espacio a las otras muestras de la clase [21].

Para cada algoritmo, se analizarán las métricas de evaluación para saber si nuestro algoritmo funciona y si el conjunto de datos que se ha obtenido contiene información relevante para determinar si un perfil puede inducir a sufrir un trastorno de la conducta alimentaria. Para todos los algoritmos, se dividirá el conjunto de datos en dos partes, una para entrenar el algoritmo, y otra para hacer tests de los datos, esto es crucial ya que si se entrena con los mismos datos que se hace test, no se podría saber si los resultados correctos son porque está sobreentrenado. La división del

conjunto que se ha escogido ha sido de 70% para entrenar y 30% para clasificar, generados de manera aleatoria. También para cada algoritmo se puede obtener el listado de características más relevantes. Con esta división disponemos de 388 perfiles para entrenar los algoritmos, y 167 para clasificar. De las 388 muestras para entrenar, 135 corresponden a perfiles no dañinos, y 253 a perfiles dañinos, es por esto que al conjunto de datos para entrenar se le ha aplicado la técnica SMOTE mencionada anteriormente, con la que se balancean los datos y se obtienen 253 perfiles tanto dañinos como no dañinos.

El primer algoritmo de aprendizaje automático utilizado ha sido el de Regresión Logística, esta técnica se utiliza para predecir resultados de clases binarias en función de una serie de características, en este caso, como se quiere predecir si un perfil es o no dañino, es un problema de predicción de clases binarias [22], en este caso si el perfil es dañino o no, se pueden observar los resultados del algoritmo en la Tabla 2, el macro de F1-Score ha sido del 60%.

	Precisión	Sensibilidad	*F1-score*	*Support*
0	0.47	0.77	0.58	60
1	0.80	0.51	0.62	107
Exactitud			0.60	167
Media macro	0.63	0.64	0.60	167
Media ponderada	0.68	0.60	0.61	167

Tabla 2: Resultados del algoritmo de Regresión Logística

Para el algoritmo de Regresión Logística, se obtienen los siguientes valores de la importancia de las características.

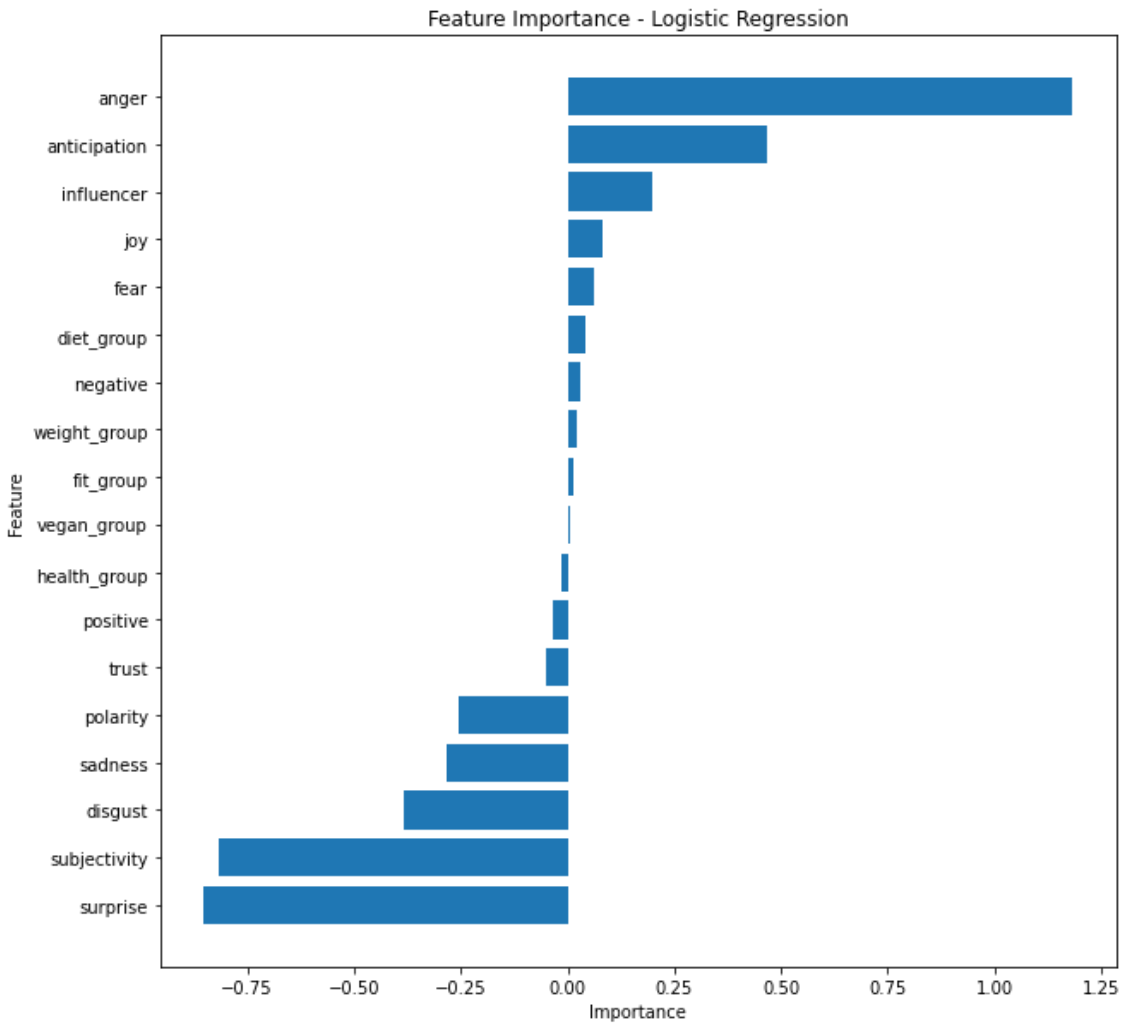

Figura 16: Importancia de las características para el algoritmo de Regresión Logística

El siguiente algoritmo utilizado fue un Árbol de Decisión, este algoritmo crea un árbol en el que ubica las características más importantes arriba, y entonces va recorriendo el árbol hasta llegar a un nodo hoja con el resultado de la clasificación [23], para este algoritmo, los resultados obtenidos son similares a los anteriores, en la Tabla 3 se encuentran los resultados y el macro de F1-Score es del 62%.

	Precisión	Sensibilidad	*F1-score*	*Support*
0	0.47	0.53	0.50	60
1	0.72	0.66	0.69	107
Exactitud			0.62	167
Media macro	0.59	0.60	0.59	167
Media ponderada	0.63	0.62	0.62	167

Tabla 3: Resultados del algoritmo de Árbol de Decisión

Para el Árbol de Decisión, la relevancia de las características que se obtiene es la siguiente:

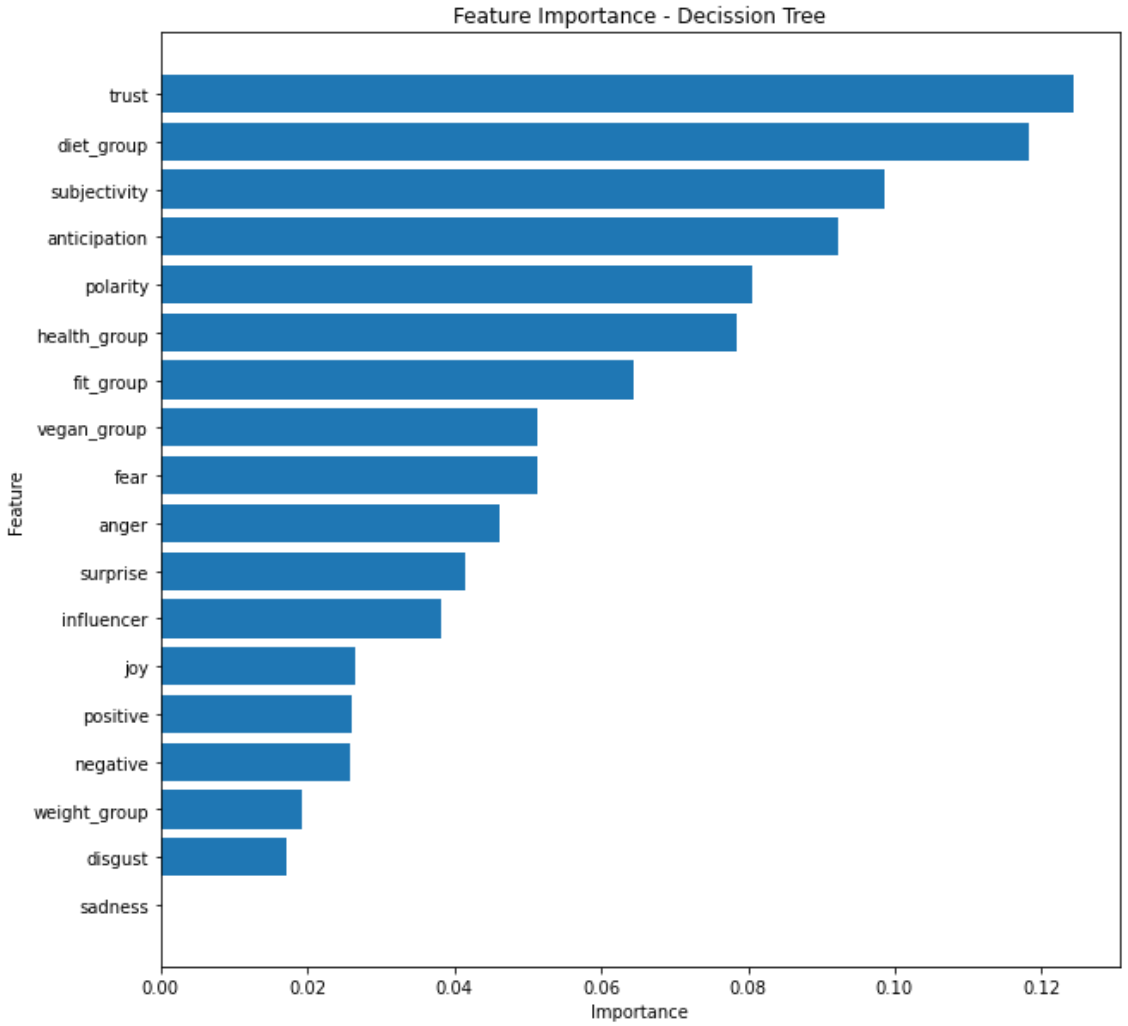

Figura 17: Importancia de las características para el algoritmo de Árbol de Decisión

Por último, se utilizó un algoritmo de *Random Forest*, este algoritmo es la unión de diferentes árboles de decisión, en la Tabla 4 se pueden observar los resultados, para este algoritmo se ha obtenido un macro de F1-score del 70%.

	Precisión	Sensibilidad	*F1-score*	*Support*
0	0.50	0.52	0.51	60
1	0.72	0.71	0.72	107
Exactitud			0.64	167
Media macro	0.61	0.61	0.61	167
Media ponderada	0.64	0.64	0.64	167

Tabla 4: Resultados del algoritmo de *Random Forest*

Para el algoritmo de *Random Forest*, la relevancia de las características es la siguiente:

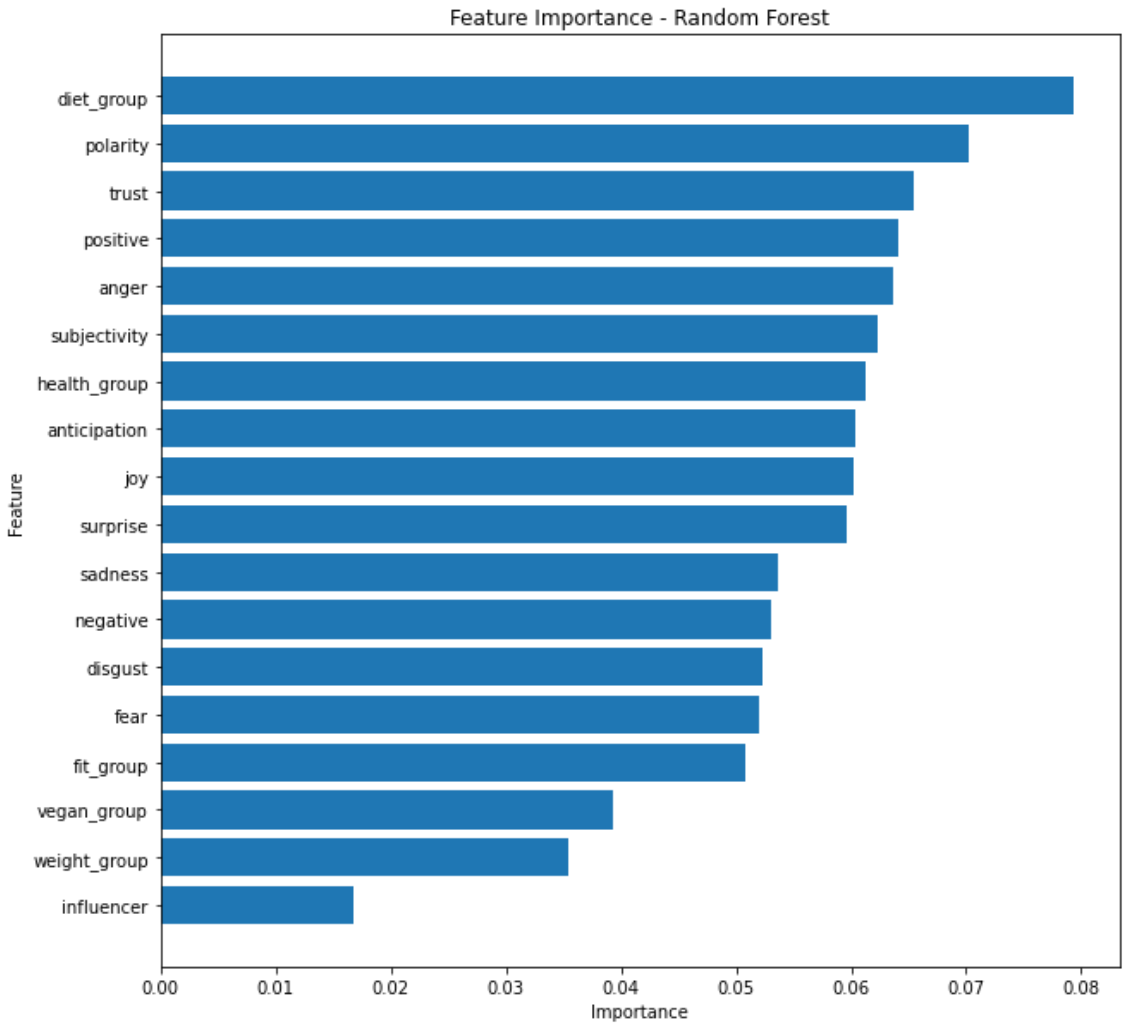

Figura 18: Importancia de las características para el algoritmo de *Random Forest*

6. Discusión de resultados

Después del análisis de los datos, se comprueba como las características más importantes para cada clasificador, a pesar de no ser las mismas, algunas se mantienen comunes en las partes superiores, como pueden ser los sentimiento de enfado y anticipación. También se observa como el grupo de *hashtags* de dieta es de las características más relevantes para los tres modelos, esto podemos relacionarlo con el gráfico visto anteriormente (Figura 15) en el que se podía observar cómo este grupo de *hashtags* era más utilizado por perfiles dañinos.

Se puede observar la importancia de la categoría de los *hashtags* bajo los que publican contenido, vemos que los perfiles peligrosos sobre todo publican fotografías con *hashtags* relacionados con dietas y estar en forma.

Los tres algoritmos tienen un macro de F1-Score de 60%. Si nos fijamos en el detalle, se observa para el algoritmo de Regresión Logística un 60% de F1-Score tanto para perfiles dañinos como no dañinos, y para los algoritmos de Árbol de Decisión y *Random Forest* se observa que para los perfiles dañinos se obtiene sobre un 70% de F1-Score y para los perfiles no dañinos sobre 50%. Esto no resulta un dato muy preocupante ya que con los datos obtenidos, los algoritmos clasifican mejor los perfiles dañinos, que es el objetivo de este proyecto. La clasificación errónea de perfiles no dañinos como perfiles dañinos son falsos positivos, pero lo perjudicial sería tener falsos negativos, y que el algoritmo clasificase perfiles que son dañinos como no dañinos.

7. Planificación del proyecto

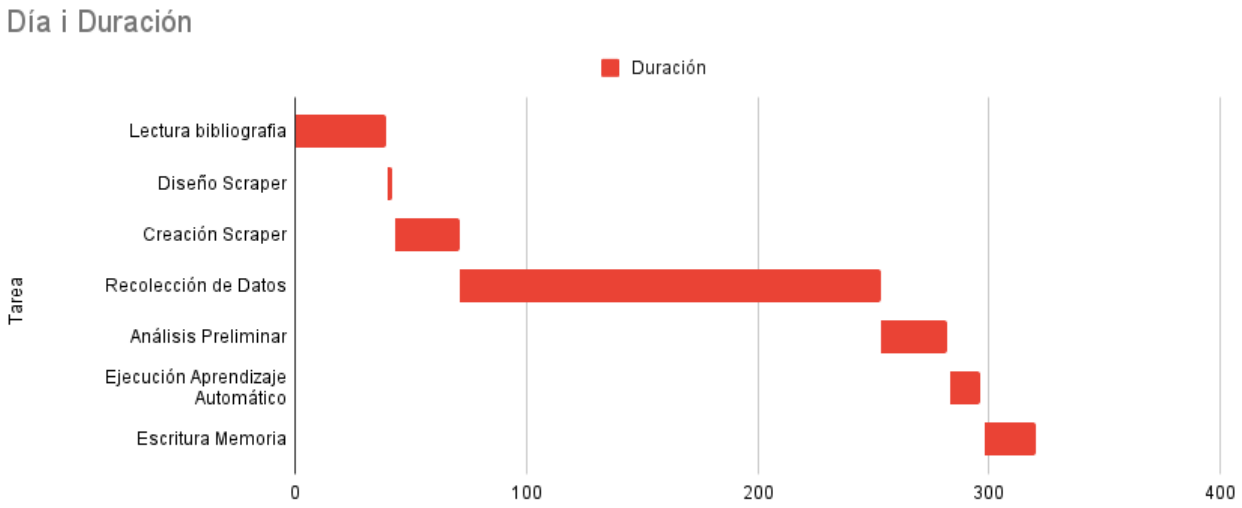

Figura 19: Diagrama de Gantt con la planificación del proyecto

La parte principal del proyecto es la recolección de datos, es por esto que las tareas a las que se le ha dedicado más tiempo han sido a la de la creación del *scraper* y la recolección y análisis de datos.

Era muy importante entender las tendencias de este tipo de publicaciones en Instagram y saber qué *hashtags* se utilizan, es por esto que en la recolección de datos también se incluye un previo análisis de diferentes *hashtags* para ver qué tipo de contenido se publicaba.

La recolección de datos es la tarea que más tiempo ha llevado debido a las restricciones de Instagram y la imposibilidad de obtener los datos por “fuerza bruta”. Para realizar esta tarea, ha sido necesario contar con conexión a internet y la capacidad de poder renovar la IP.

8. Conclusiones y trabajo futuro

Como se ha podido evidenciar, existe un conjunto de cuentas en Instagram que bajo el pretexto de vivir una vida más saludable, comer sano y hacer deporte, pueden llegar a producir una imagen distorsionada de la realidad para determinados colectivos que están en riesgo de sufrir, o están sufriendo, un TCA, esto es debido a que aunque muchas de estas tendencias de Instagram se expongan como algo inofensivo y que las personas están en pleno control, las personas que hace estas publicaciones no son conocedoras del alcance real de sus acciones, ni de a qué tipo de colectivos les llegan sus publicaciones. Es por este motivo, que este proyecto ha evidenciado que es posible crear un clasificador de publicaciones que evite que personas que se encuentran en una situación vulnerable por estar sufriendo un TCA se vean influenciadas negativamente e incluso empeorada su condición.

De cara a futuros trabajos, vemos como un clasificador de estas características puede extenderse a diferentes redes sociales. Es muy importante entender las diferentes tendencias de redes sociales entre la población, ya que muchas veces cuando aparece una nueva y no se controla, pueden llegarse a crear comunidades peligrosas para determinados colectivos vulnerables. También, viendo que este clasificador ha obtenido buenos resultados clasificando los perfiles peligrosos, pero el porcentaje de precisión ha sido inferior para los perfiles no peligrosos, en futuros trabajos se podría intentar obtener más muestras de perfiles no dañinos, o investigar features que proporcionen más diferencia entre perfiles dañinos y no dañinos.

Es de especial importancia entender que Instagram es una red social poco permisiva, y que publicaciones que inciten a un TCA de manera más explícita son denunciadas y eliminadas, es por esto que encontrar estos resultados, y ver que perfiles que siguen a cuentas aparentemente inofensivas están en riesgo de tener un TCA es alarmante. Con esto se evidencia cómo hay que prestar especial atención a redes sociales donde esto no sucede, en redes sociales como Twitter y Tumblr la censura es mínima, y pueden llegar a crearse comunidades muy peligrosas que si una persona en una situación vulnerable las encuentra puede resultar muy dañino. Es por eso que analizar y vigilar todas estas redes sociales no puede ser visto desde el punto de vista de censurar, sino de proteger.

Bibliografía

[1] National Institute of Mental Health. (s.f.). *Eating Disorders: About More Than Food*. Recuperado el 2 de Mayo de 2021, de National Institute of Mental Health (NIM): https://www.nimh.nih.gov/health/publications/eating-disorders/

[2] Mouzo, J. (01 de Diciembre de 2020). *La pandemia agudiza los trastornos de la conducta alimentaria*. Recuperado el 16 de Mayo de 2021, de El País: https://elpais.com/sociedad/2020-11-30/la-pandemia-agudiza-los-trastornos-de-la-conducta-alimentaria.html

[3] Asociación TCA Aragón. (01 de Junio de 2020). *Estadísticas sobre los TCA*. Recuperado el 16 de Mayo de 2021, de Asociación TCA Aragón: https://www.tca-aragon.org/2020/06/01/estadisticas-sobre-los-tca/

[4] National Eating Disorders Association (NEDA). (s.f.). *Media & Eating Disorders*. Recuperado el 18 de Mayo de 2021, de National Eating Disorders Association (NEDA): https://www.nationaleatingdisorders.org/media-eating-disorders

[5] Gleissner, G. (05 de Octubre de 2017). *Social Media and its Effect on Eating Disorders*. Recuperado el 16 de Mayo de 2021, de Huffpost: https://www.huffpost.com/entry/social-media-and-its-effect-on-eating-disorders_b_591343bce4b0e3bb894d5caa

[6] Zhou, S., Zhao, Y., Rizvi, R., Bian, J., Haynos, A. F., & Zhang, R. (2019). *Analysis of Twitter to Identify Topics Related to Eating Disorder Symptoms.*

[7] He, L., & Luo, J. (2016). *“What makes a pro eating disorder hashtag”: Using hashtags to identify pro eating disorder tumblr posts and Twitter users.*

[8] Holland, G., & Tiggemann, M. (2016). *“Strong beats skinny every time”: Disordered eating and compulsive exercise in women who post fitspiration on Instagram.*

[9] Turner, P. G., & Lefevre, C. E. (2017). *Instagram use is linked to increased symptoms of orthorexia nervosa.*

[10] Carrotte, E., Vella, A. M., & Lim, M. S. (2015). *Predictors of "Liking" Three Types of Health and Fitness-Related Content on Social Media: A Cross-Sectional Study.*

[11] Barr, S. (12 de Diciembre de 2018). *Instagram is clamping down on hashtags referencing eating disorders*. Recuperado el 22 de Mayo de 2021, de Independent: https://www.independent.co.uk/life-style/health-and-families/instagram-eating-disorder-hashtag-body-mental-health-social-media-bbc-investigation-a8679106.html

[12] Geeks for Geeks. (22 de Junio de 2020). *What is Web Scraping and How to Use It?* Recuperado el 14 de Mayo de 2021, de Geeks for Geeks: https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/

[13] BBVA. (08 de Noviembre de 2019). *'Machine learning': ¿qué es y cómo funciona?* Recuperado el 16 de Mayo de 2021, de BBVA: https://www.bbva.com/es/machine-learning-que-es-y-como-funciona/

[14] Delua, J. (12 de Marzo de 2021). *Supervised vs. Unsupervised Learning: What's the Difference?* Recuperado el 03 de Junio de 2021, de IBM: https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning

[15] IBM Cloud Education. (15 de Julio de 2020). *Machine Learning*. Recuperado el 05 de Junio de 2021, de IBM: https://www.ibm.com/cloud/learn/machine-learning

[16] Rafie, M. (21 de Marzo de 2020). *Machine Learning and IBM Watson Studio*. Recuperado el 05 de Junio de 2021, de IBM: https://developer.ibm.com/recipes/tutorials/machine-learning-and-ibm-watson-studio/

[17] Scikit Learn. (s.f.). *Decision Trees*. Recuperado el 16 de Junio de 2021, de Scikit Learn: https://scikit-learn.org/stable/modules/tree.html

[18] Scikit Learn. (s.f.). *Random Forest Classifier*. Recuperado el 05 de Junio de 2021, de Scikit Learn:

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[19] Scikit Learn. (s.f.). *Precision Recall F-score Support*. Recuperado el 05 de Junio de 2021, de Scikit Learn: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html

[20] Perez, V. (2020). *Forecasting Apple Inc Volatility Shares using Twitter Sentiment Analysis.*

[21] Brownlee, J. (2020). *SMOTE for Imbalanced Classification with Python.*

[22] Li, S. (29 de Septiembre de 2017). *Building A Logistic Regression in Python, Step by Step*. Recuperado el 16 de Mayo de 2021, de Towards Data Science: https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8

[23] Robinson, S. (02 de Febrero de 2018). *Decision Trees in Python with Scikit-Learn*. Recuperado el 14 de Junio de 2021, de Stack Abuse: https://stackabuse.com/decision-trees-in-python-with-scikit-learn