Análisis y Cuantificación del Uso de Datos Sensibles por Parte de Facebook

José González Cabañas, Ángel Cuevas, and Rubén Cuevas
Departmento de Ingeniería Telemática
Universidad Carlos III de Madrid
{jgcabana, acrumin, rcuevas}@it.uc3m.es

Resumen

El reciente Reglamento General de Protección de Datos (RGPD) de la Unión Europea restringe el procesamiento y la explotación de algunas categorías de datos personales (salud, orientación política, preferencias sexuales, creencias religiosas, origen étnico, etc.) debido a los riesgos de privacidad que pueden resultar del uso malicioso de dicha información. El RGPD se refiere a estas categorías como datos personales sensibles. Este artículo cuantifica la proporción de usuarios de Facebook en la Unión Europea (UE) que fueron etiquetados con intereses relacionados a datos personales potencialmente sensibles en el periodo anterior a que el RGPD entrase en vigor. Los resultados de nuestro estudio sugieren que Facebook etiqueta al 73% de usuarios de la UE con intereses potencialmente sensibles. Esto se corresponde con un 40% del total de ciudadanos de la UE. También, hemos estimado que un atacante malicioso podría desvelar la identidad de usuarios de Facebook que han sido asignados con un interés potencialmente sensible a un coste bajo de 0.015€ por usuario. Finalmente, proponemos e implementamos una extensión web que informe a los usuarios de Facebook del los intereses potencialmente sensibles que Facebook les ha asignado.

1 Introducción

Los ciudadanos de la Unión Europea (UE) han demostrado serias preocupaciones con respecto al tratamiento de la información personal por parte de los servicios online. El Eurobarómetro de 2015 sobre protección de datos [21] revela que: el 63% de los usuarios de la UE no confía en los negocios online, a más de la mitad no les gusta proporcionar información personal a cambio de servicios gratuitos y el 53% no quiere que compañías de Internet usen su información personal en publicidad personalizada. La UE ha reaccionado a las preocupaciones de los ciudadanos con la aprobación del Reglamento General de Protección de Datos (RGPD) [8], que define un nuevo marco regulatorio para el tratamiento de la información personal. Los Estados Miembros de la UE tuvieron hasta Mayo de 2018 para incorporar este reglamento en su legislación nacional.

El RGPD (y anteriores leyes nacionales de protección de datos en la UE) define algunas categorías de datos personales como sensibles y prohíbe el procesamiento de ellos salvo algunas limitadas excepciones (por ejemplo, que el usuario dé un consentimiento explícito para el procesamiento de esta información con una finalidad específica). Estas categorías de datos son definidas como *"Datos Especialmente Protegidos"*, *"Categorías Especiales de Datos Personales"* o *"Datos Sensibles"*. En particular, el RGDP define como datos personales sensibles: *"datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales de una persona física"*.

Debido a las implicaciones legales, éticas y de privacidad sobre el tratamiento de datos personales sensibles, es importante conocer si los servicios en línea están explotando comercialmente esta información sensible. Si es el caso, también es esencial medir la proporción de usuarios/ciudadanos que pueden estar afectados por la explotación de sus datos personales sensibles. En este artículo enfocamos estas preguntas centrándonos en la *publicidad online* que representa la mayor fuente de beneficios para la mayoría de los servicios online. En particular, nos centramos en Facebook (FB), cuya plataforma de publicidad online es la segunda en términos de ganancias únicamente por detrás de Google [2].

Facebook etiqueta usuarios con las denominadas preferencias de publicidad (ad preferences), que representan intereses potenciales de los usuarios. FB asigna a

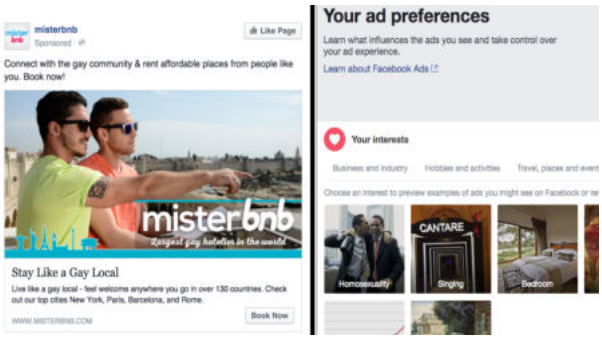


Figure 1: Captura de pantalla de un anuncio recibido por uno de los autores de este artículo y la lista de preferencias de publicidad mostrando que FB infirió que dicha persona estaba interesada en *Homosexualidad*.

los usuarios diferentes preferencias publicitarias basándose en su actividad en línea tanto en esta red social como en webs de terceros que son monitorizadas por FB. Los anunciantes que realizan campañas publicitarias pueden llegar a grupos de usuarios que han sido etiquetados con cierta preferencia publicitaria (e.g., llegar a usuarios de FB interesados en *"Starbucks"*). Algunas de estas preferencias publicitarias indican opiniones políticas, orientación sexual, salud y otros atributos potencialmente sensibles. De hecho, uno de los autores de este artículo recibió el anuncio mostrado en la Figura 1 (izquierda). El autor no había definido explícitamente su orientación sexual pero descubrió que FB le había asignado la preferencia publicitaria *"Homosexualidad"* (ver Figura 1 derecha). Nuestros datos sugieren que similares asignaciones de preferencias publicitarias potencialmente sensibles ocurren mucho más a menudo. Por ejemplo, páginas web finales asociadas a anuncios mostrados a usuarios de FB incluidos en nuestro estudio incluyen: *iboesterreich.at* (política), *gaydominante.com* (sexualidad), *elpartoestuyo.com* (salud).

Esto ilustra que FB puede estar de hecho procesando información personal sensible, algo que está ahora prohibido bajo el RGPD de la UE sin el consentimiento explícito del usuario y también anteriormente prohibido por reglamentos nacionales de protección de datos en Europa. Recientemente, la Agencia Española de Protección de Datos (AEPD) multó a FB con 1.2M€ por violar la regulación española de protección de datos [6]. La AEPD argumentó que FB *"Almacena, guarda y usa datos, incluidos datos especialmente protegidos, con fines publicitarios sin obtener consentimiento"*.

Motivados por estos eventos y la promulgación del RGDP en la Unión Europea, este documento examina el uso de datos potencialmente sensibles de Facebook hasta enero de 2018, solo unos meses antes de que el RGPD entrara en vigor. El objetivo principal de este documento es *cuantificar la proporción de ciudadanos de la UE y usuarios de FB a los que se les pueden haber asignado preferencias de anuncio vinculadas a datos personales potencialmente sensibles*. Dejamos el análisis de las prácticas de datos de Facebook después de la fecha de vigencia del RGPD del 25 de mayo de 2018 (cuando las violaciones podrían ser exigibles) para trabajos futuros.

Para lograr nuestro objetivo, analizamos más de 5.5M de preferencias publicitarias (126K únicas) asignadas a más de 4.5K usuarios de FB que han instalado la extensión de navegador Data Valuation Tool for Facebook Users (FDVT) [12]. La razón para usar las preferencias publicitarias asignadas a los usuarios del FDVT es que podemos probar que las preferencias publicitarias consideradas en nuestro estudio han sido efectivamente asignadas a usuarios reales.

La primera contribución de este documento es una metodología que combina técnicas de procesamiento de lenguaje natural y clasificación manual realizada por 12 miembros evaluadores para obtener las preferencias publicitarias en nuestro conjunto de datos potencialmente vinculados a datos personales sensibles. Estas preferencias publicitarias pueden desvelar: origen étnico o racial, opiniones políticas, creencias religiosas, información de salud u orientación sexual. Por ejemplo, las preferencias publicitarias *"Homosexualidad"* y *"Comunismo"* pueden revelar la orientación sexual y la preferencia política de un usuario, respectivamente.

Una vez que hemos identificado la lista de preferencias publicitarias potencialmente sensibles, la usamos para consultar el Administrador de Anuncios de FB y obtener el número de usuarios y ciudadanos de FB expuestos a estas preferencias publicitarias en toda la UE, así como en cada uno de sus Estados Miembros. Esta cuantificación es nuestra segunda contribución, que cumple con el objetivo principal del artículo.

Finalmente, después de ilustrar los riesgos de éticos y de privacidad derivados de la explotación de estas preferencias publicitarias de FB, presentamos una ampliación del FDVT que informa a los usuarios de las preferencias publicitarias potencialmente sensibles que FB les ha asignado. Esta es la última contribución de este artículo.

Nuestra investigación nos lleva a las siguientes conclusiones principales:

- **Hemos identificado 2092 (1.66%) preferencias publicitarias potencialmente sensibles de las 126K presentes en nuestro conjunto de datos.**

- **FB asigna en promedio 16 preferencias publicitarias potencialmente sensibles para los usuarios del FDVT.**

- **Más del 73% de los usuarios de FB de la UE, que corresponde al 40% de ciudadanos de la UE, es-**

tán etiquetados con al menos una de las 500 principales (es decir, las más populares) preferencias publicitarias potencialmente sensibles de nuestro conjunto de datos.

- **Las mujeres tienen una exposición significativamente mayor que los hombres a preferencias publicitarias potencialmente sensibles. De manera similar, el grupo de edad adulta temprana (20-39 años) tiene la mayor exposición de todos los grupos de edad.**

- **Realizamos una estimación aproximada que sugiere que desvelar la identidad de los usuarios de FB etiquetados con preferencias publicitarias potencialmente sensibles puede ser tan barato como 0.015€ por usuario.**

2 Fundamentos

2.1 Facebook Ads Manager

Los anunciantes configuran sus campañas de anuncios a través del Administrador de Anuncios de Facebook (FB).[1] Este, permite a los anunciantes definir la audiencia (es decir, el perfil de usuario) a la que desean llegar con sus campañas publicitarias. Se puede acceder a través de un panel de control o una API. FB Ads Manager ofrece a los anunciantes una amplia gama de parámetros de configuración como (pero no limitado a): *ubicación* (país, región, ciudad, código postal, etc.), *parámetros demográficos* (género, edad , idioma, etc.), *hábitos* (dispositivo móvil, sistema operativo y/o navegador web utilizado, frecuencia de viaje, etc.), e *intereses* (deportes, comida, automóviles, belleza, etc.).

El parámetro *interés* es el más relevante para nuestro trabajo. Incluye cientos de miles de posibilidades que capturan el interés de los usuarios de cualquier tipo. Estos intereses están organizados en una estructura jerárquica con varios niveles. El primer nivel está formado por 14 categorías.[2] Además de los intereses incluidos en esta jerarquía, el Administrador de Anuncios de FB ofrece una barra de búsqueda *"Detail Targeting"* donde los usuarios pueden escribir cualquier texto libre y sugiere intereses vinculados a tal texto. En este documento, aprovechamos el parámetro *interés* para identificar posibles intereses sensibles.

Los anunciantes pueden configurar sus audiencias objetivo en función de cualquier combinación de los parámetros descritos. Un ejemplo de una audiencia podría ser *"Usuarios que viven en Italia, de entre 30 y 40 años, hombres e interesados en comida rápida"*.

Finalmente, FB Ads Manager proporciona información detallada sobre la audiencia configurada. El parámetro más relevante para nuestro artículo es el *Alcance potencial* que informa del número de usuarios de FB registrados que coinciden con la audiencia definida.

2.2 Preferencias publicitarias de Facebook

FB asigna a cada usuario un conjunto de preferencias publicitarias, es decir, un conjunto de intereses, derivados de los datos y la actividad del usuario en FB y sitios web externos, aplicaciones y servicios en línea donde FB está presente. Estas preferencias publicitarias son, de hecho, los intereses que se ofrecen a los anunciantes en el FB Ads Manager para configurar sus audiencias.[3] Por lo tanto, si se asigna a un usuario *"Relojes"* dentro de su lista de preferencias publicitarias, será un objetivo potencial de cualquier campaña publicitaria de FB configurada para llegar a los usuarios interesados en relojes.

Cualquier usuario puede acceder y editar (agregar o eliminar) sus preferencias publicitarias,[4] pero sospechamos que hay pocos usuarios conscientes de esta opción. Cuando un usuario coloca el ratón sobre una preferencia de anuncio específico, una ventana emergente indica por qué al usuario se le ha asignado esta preferencia de anuncio. Al examinar 5.5M preferencias publicitarias asignadas a usuarios de FDVT (ver Subsección 2.3), hemos encontrado 6 razones para la asignación de preferencias publicitarias: (i)*Esta es una preferencia que añadió*, (ii) *Tienes esta preferencia porque creemos que puede ser relevante para ti según lo que haces en Facebook, como las páginas que te han gustado o los anuncios en los que has hecho clic*, (iii) *Tienes esta preferencia porque hiciste clic en un anuncio relacionado con ...*, (iv) *Tienes esta preferencia porque instalaste la aplicación ...*, (v) *Tienes esta preferencia porque te gustó una página relacionada con ...*, (vi) *Tienes esta preferencia debido a comentarios, publicaciones, acciones o reacciones que hiciste relacionadas con ....*

2.3 FDVT

Data Valuation Tool for Facebook Users (FDVT) [12] es una extensión de navegador actualmente disponible para

[1] `https://www.facebook.com/ads/manager`

[2] Negocios e industria, Educación, Familia y relaciones, Fitness y bienestar, Alimentos y bebidas, Pasatiempos y actividades, Estilo de vida y cultura, Noticias y entretenimiento, Gente, Compras y moda, Deportes y al aire libre, Tecnología, Lugares y eventos de viaje, Vacío.

[3] Dado que los intereses y las preferencias publicitarias se refieren a la misma cosa, usamos estos dos términos indistintamente en el resto del artículo

[4] Acceder y editar la lista de preferencias publicitarias: `https://facebook.com/ads/preferences/edit`

Google Chrome[5] y Mozilla Firefox.[6] Proporciona a los usuarios de FB una estimación en tiempo real de los ingresos que se están generando para Facebook según su perfil y la cantidad de anuncios que ven y hacen clic durante una sesión de Facebook. Más de 6K usuarios han instalado el FDVT desde su lanzamiento público en octubre de 2016 y febrero de 2018. El FDVT recopila (entre otros datos) las preferencias publicitarias que FB asigna al usuario. Aprovechamos esta información para identificar las preferencias publicitarias potencialmente sensibles asignadas a los usuarios que han instalado el FDVT.

3 Consideraciones legales

3.1 Reglamento General de Protección de Datos

El Reglamento General de Protección de Datos de la UE (RGPD) [8] entró en vigor en mayo de 2018 y es el reglamento de referencia de protección de datos en los 28 países de la UE. El RGPD incluye un artículo que regula el uso de *Datos personales sensibles*. El artículo 9 se titula *"Procesamiento de categorías especiales de datos personales"* y establece en su primer párrafo: *"Quedan prohibidos el tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales de una persona física."*.

Después de enumerar estas prohibiciones particulares, el RGPD introduce diez excepciones (ver Apéndice A) para las cuales no se aplicará el párrafo 1 del artículo. Según nuestro conocimiento, ninguna de estas excepciones para el procesamiento de datos personales sensibles parece aplicarse al caso de las preferencias publicitarias de FB. Por lo tanto, etiquetar a los usuarios de FB con preferencias publicitarias asociadas con datos personales sensibles puede contravenir el artículo 9 de la GDPR.

3.2 Facebook multado en España

En septiembre de 2017, la Agencia Española de Protección de Datos (AEPD) multó a Facebook con 1.2M€ por violar la implementación española de la Directiva de Protección de Datos de la UE 95/46EC [1] antes del RGPD. En la resolución de la multa [6], la AEPD afirma que FB recopila, almacena y procesa datos personales protegidos con fines publicitarios sin obtener el consentimiento de los usuarios. Más detalles sobre la resolución AEPD se proporcionan en el Apéndice B.

La AEPD afirma que el uso de datos sensibles con fines publicitarios a través de la asignación de preferencias publicitarias a los usuarios por parte de FB violó la Regulación Española de Protección de Datos (y quizás las regulaciones de otros estados miembros de la UE que implementaban en sus leyes nacionales la Directiva 95/46EC de protección de datos de la UE[1], recientemente reemplazada por el RGPD).

3.3 Términos de uso de Facebook

Hemos revisado cuidadosamente los términos y políticas de FB. Aunque no somos abogados, no encontramos una revelación clara a los usuarios de la UE de que FB procesa y almacena datos personales sensibles específicamente, ni un lugar donde los usuarios pueden dar su consentimiento. Según nuestro conocimiento, ambos son requeridos por el RGPD. Además, no hemos encontrado ninguna prohibición general por parte de FB en los anunciantes que buscan orientar anuncios basados en datos personales sensibles. Se proporcionan más detalles sobre el análisis de los términos de servicio de FB en el Apéndice C.

4 Conjunto de datos

Para descubrir preferencias publicitarias potencialmente sensibles y cuantificar cuántas son las cuentas de usuario de FB de la UE asociadas a ellas, buscamos recopilar un conjunto de datos de preferencias publicitarias vinculadas a las cuentas reales de FB de la UE. Si detectamos preferencias publicitarias que representan datos personales potencialmente sensibles, este conjunto de datos proporcionará evidencia de que las preferencias están asignadas a cuentas reales de FB. En función de este objetivo, nuestro conjunto de datos se crea a partir de las preferencias publicitarias recopiladas de usuarios reales que han instalado el FDVT. Observamos que la cantidad de preferencias publicitarias recuperadas por el FDVT representa solo un subconjunto del conjunto general de preferencias, pero podemos garantizar que se asignaron a cuentas reales. Nuestro conjunto de datos incluye las preferencias publicitarias de 4577 usuarios que instalaron el FDVT entre octubre de 2016 y octubre de 2017, de los cuales 3166 son de algún país de la UE. A estos 4577 usuarios del FDVT se les han asignado 5,5M de preferencias publicitarias, de las cuales 126192 son únicas.

[5] https://chrome.google.com/webstore/detail/fdvt-social-network-data/blednbbpnnambjaefhlocghajeohlhmh

[6] https://addons.mozilla.org/firefox/addon/fdvt

Nuestro conjunto de datos incluye la siguiente información para cada preferencia de anuncio:

-*ID de la preferencia de anuncio*: esta es la clave que utilizamos para identificar una preferencia de anuncio independientemente del idioma utilizado por un usuario de FB. Por ejemplo, a la preferencia de anuncio *Milk, Leche, Lait* que se refiere a lo mismo en inglés, español y francés, se le asigna un solo ID de FB. Por lo tanto, podemos identificar de forma única cada preferencia de anuncio en todos los países e idiomas de la UE.

-*Nombre de la preferencia de anuncio*: Este es el descriptor principal de la preferencia de anuncio. FB devuelve una versión unificada del nombre para cada ID de preferencia de anuncio, generalmente en inglés. Por lo tanto, tenemos el nombre en inglés de las preferencias publicitarias independientemente del idioma original en la colección de los datos. Notamos que, en algunos casos, traducir el nombre de preferencia del anuncio no tiene sentido (por ejemplo, el caso de los nombres de personas: celebridades, políticos, etc.).

-*Categoría de desambiguación*: para algunas preferencias publicitarias, Facebook agrega esto en un campo separado o entre paréntesis para aclarar el significado de una preferencia de anuncio en particular que puede tener varios significados (por ejemplo, violeta (color); violeta: ropa (marca)). Hemos identificado más de 700 categorías de desambiguación diferentes (por ejemplo, ideología política, enfermedad, libro, sitio web, equipo deportivo, etc.). Entre las 126K preferencias publicitarias analizadas, el 87% incluye este campo.

-*Categoría*: en muchos casos, algunas de las 14 categorías de primer nivel introducidas en la Sección 2.1 se asignan para contextualizar las preferencias publicitarias. Por ejemplo, Manchester United F.C. Está vinculado a los deportes y al aire libre.

-*Tamaño de la audiencia*: este valor informa del número de usuarios de Facebook a los que se les ha asignado la preferencia de anuncio en todo el mundo.

-*Motivo por el cual se agregó al usuario*: El motivo por el cual la preferencia de anuncio se ha asignado al usuario de acuerdo con FB. Hay seis razones posibles introducidas en la subsección 2.2.

La Figura 2 muestra la CDF del número de preferencias publicitarias por usuario. A cada usuario del FDVT se le asigna una mediana de 474 preferencias. Además, la Figura 3 muestra la CDF de la proporción de usuarios del FDVT (eje x) a los que se asignó una preferencia de anuncio determinada (eje y). Observamos una distribución muy sesgada que indica que la mayoría de las preferencias publicitarias están asignadas a una pequeña fracción de usuarios. Por ejemplo, cada preferencia de anuncio se asigna a una mediana de sólo 3 (0.06%) usuarios del FDVT. Sin embargo, es importante tener en cuenta que muchas preferencias publicitarias todavía llegan a una parte razonable de los usuarios. Nuestro conjunto de datos incluye 1000 preferencias publicitarias que alcanzan al menos el 11% de los usuarios del FDVT.

5 Metodología

Buscamos cuantificar el número de usuarios de FB de la UE a los que se les han asignado preferencias publicitarias potencialmente sensibles. Para este fin, utilizamos las 126K preferencias publicitarias únicas asignadas a los usuarios del FDVT y seguimos un proceso de dos pasos. En el primer paso, combinamos las técnicas de Procesamiento de Lenguaje Natural (PLN) con la clasificación manual para obtener una lista de las 126K preferencias publicitarias probablemente sensibles consideradAs. En el segundo paso, aprovechamos la API de FB Ads Manager para cuantificar a cuántos usuarios de FB en cada país de la UE se les ha asignado al menos una de las preferencias publicitarias etiquetadas como potencialmente sensibles.

5.1 Identificación de preferencias publicitarias potencialmente sensibles

Contamos con un grupo de investigadores con cierto conocimiento en el área de privacidad para identificar manualmente las preferencias publicitarias potencialmente sensibles dentro de nuestro grupo de 126K preferencias publicitarias recuperadas de los usuarios del FDVT. Sin embargo, la clasificación manual de las 126K preferencias publicitarias no sería factible.[7].

5.1.1 Pre-filtrado

Categorías sensibles: Para identificar las preferencias publicitarias probablemente sensibles de manera automatizada, seleccionamos cinco de las categorías relevantes enumeradas como *Datos personales sensibles* por el RGPD: (i) datos que revelan el origen racial o étnico, (ii) datos que revelan opiniones políticas, (iii) datos que revelan creencias religiosas o filosóficas, (iv) datos sobre salud y (v) datos sobre vida sexual y orientación sexual. Seleccionamos estas categorías porque una inspección manual preliminar indicó que existen preferencias publicitarias en nuestro conjunto de datos que probablemente pueden revelar información relacionada con

[7]Si consideramos 10s como el tiempo promedio requerido para clasificar una preferencia de anuncios como sensible frente a no sensible, esta tarea requeriría 44 días completos de ocho horas. Esta tarea de clasificación manual no es escalable por lo que aprovechamos las técnicas de PLN para filtrar previamente la lista de preferencias publicitarias que probablemente sean más sensibles. Esta fase de filtración entregará un subconjunto de las preferencias publicitarias probablemente sensibles que se pueden clasificar manualmente en un tiempo razonable

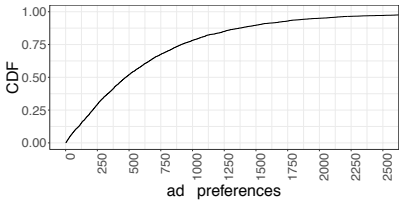


Figure 2: CDF del número de preferencias de anuncio por usuario en el FDVT.

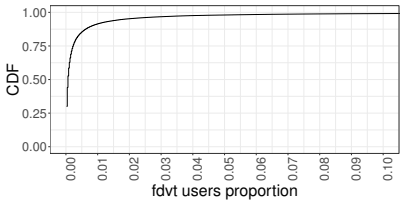


Figure 3: CDF de la proporción de usuarios del FDVT (eje x) por preferencia de anuncio (eje y).

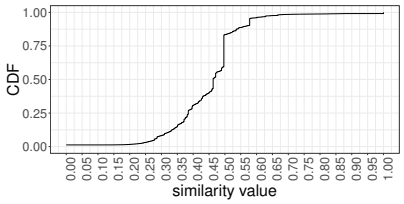


Figure 4: CDF de la puntuación de similitud semántica asociada cada una de las 126K preferencias de anuncio obtenidas del conjunto de datos del FDVT.

ellas. Por ejemplo, las preferencias publicitarias *"Socialismo"*, *"Islam"*, *"Salud reproductiva"*, *"Homosexualidad"* o *"Feminismo negro"* pueden sugerir *opinión política, creencia religiosa, problema de salud, orientación sexual* o *origen étnico o racial* de los usuarios a los que se les han asignado, respectivamente. Tenga en cuenta que todos estos ejemplos de preferencias publicitarias se han extraído de nuestro conjunto de datos; por lo tanto, han sido asignadas a usuarios reales de FB.

Nuestro proceso automatizado clasificará una preferencia de anuncio como *probablemente sensible* si podemos mapear semánticamente ese nombre de preferencia de anuncio en una de las cinco categorías sensibles analizadas en este documento. Para este fin, hemos definido un diccionario que incluye palabras clave y frases cortas representativas de cada una de las cinco categorías consideradas sensibles. Utilizamos dos fuentes de datos para crear el diccionario: primero, una lista de temas controvertidos disponibles en Wikipedia.[8] En particular, seleccionamos las siguientes categorías de esta lista: política y economía, religión y sexualidad. En segundo lugar, obtuvimos una lista de palabras con un significado semántico muy similar a las cinco categorías de datos personales sensibles. Para este fin, utilizamos la API de Datamuse,[9], un motor de consulta de búsqueda de palabras que permite a los desarrolladores encontrar palabras que coincidan con un conjunto de condiciones. Entre otras características, Datamuse permite *"encontrar palabras con un significado similar a X"* mediante una consulta simple.

El diccionario final incluye 264 palabras clave.[10] Aprovechamos las palabras clave de este diccionario para encontrar preferencias publicitarias que presentan una gran similitud semántica con al menos una palabra clave. En estos casos, los etiquetamos como preferencias publicitarias probablemente sensibles. Vale la pena señalar que este enfoque hace que nuestra metodología sea flexible, ya que el diccionario se puede ampliar para incluir nuevas palabras clave para las categorías consideradas u otras categorías, lo que puede descubrir preferencias publicitarias potencialmente sensibles adicionales.

A continuación describimos en detalle el cálculo de la similitud semántica.

Cálculo de similitud semántica: El proceso de cálculo de similitud semántica toma dos entradas: las 126K preferencias publicitarias de nuestro conjunto de datos del FDVT y el diccionario de 264 palabras clave asociado con las categorías sensibles consideradas. Calculamos la similitud semántica de cada preferencia de anuncio con todas las 264 palabras clave del diccionario. Para cada preferencia de anuncio, registramos el valor de similitud más alto de las 264 operaciones de comparación. Como resultado de este proceso, a cada una de las 126K preferencias publicitarias se le asigna una puntuación de similitud, lo que indica su probabilidad de ser una preferencia de anuncios sensible.

Para implementar la tarea de comparación de similitud semántica, aprovechamos el paquete Spacy para python[11] (consulte los detalles sobre Spacy en el Apéndice D). Elegimos Spacy porque anteriormente se ha utilizado en la literatura para fines de procesamiento de texto que ofrece un buen rendimiento [15][22]. Además, Spacy ofrece una buena escalabilidad. Calcula la similitud semántica de 33314688 ítems (126192 x 264) en 7 minutos usando un servidor con doce núcleos de 2.6 GHz y 96 GB de RAM. Para realizar nuestro análisis, aprovechamos la función *similarity* de Spacy. Esta característica permite comparar palabras, tramos de texto o documentos, y calcula la similitud semántica entre ellos. La salida es un valor de similitud semántica que oscila entre 0 y 1. Cuanto más cerca de 1, mayor es la similitud semántica.

[8] https://en.wikipedia.org/wiki/Wikipedia:List_of_controversial_issues

[9] https://www.datamuse.com/api/

[10] https://fdvt.org/usenix2018/keywords.html

[11] https://spacy.io

Este proceso reveló valores de similitud muy bajos para algunos casos en los que la preferencia de anuncios analizados coincidía con la definición de algunas de las categorías de datos personales sensibles. Algunos de estos casos son: personas físicas, como políticos (que pueden revelar la opinión política del usuario); partidos políticos con nombres que no incluyen ningún término político estándar; enfermedades de salud o lugares de cultos religiosos que pueden tener nombres con baja similitud semántica con palabras clave relacionadas con la salud y la religión en nuestro diccionario, respectivamente. Tres ejemplos que ilustran los casos referidos son: < nombre: "Angela Merkel", desambiguación: político >; < nombre: "I Love Italy", desambiguación: Partido político >; < nombre: ejercicio "Kegel", desambiguación: procedimiento médico >. En la mayoría de estos casos, la categoría de desambiguación es más útil que el nombre de preferencia de anuncios cuando se realiza el análisis de similitud semántica. Por ejemplo, en el caso de los nombres de políticos, partidos políticos y enfermedades de salud, el campo de categoría de desambiguación incluye el término *"politics"*, *"Political Party"* y *"disease"*, respectivamente. Este campo también es muy útil para determinar la definición de nombres de preferencias publicitarias que tienen varios significados.

En general, encontramos que para clasificar las preferencias publicitarias, la categoría de desambiguación, cuando está disponible, es un proxy mejor que el nombre de la preferencia de anuncio. Por lo tanto, si la preferencia de anuncio bajo análisis tiene un campo de categoría de desambiguación, usamos la cadena de categoría de desambiguación en lugar del nombre de preferencia de anuncio para obtener la puntuación de similitud semántica de la preferencia de anuncio.

Selección de preferencias publicitarias probablemente sensibles: El proceso de cálculo de similitud semántica asigna una puntuación de similitud a cada una de las 126K preferencias publicitarias en nuestro conjunto de datos. Esta puntuación de similitud representa la probabilidad anticipada de que una preferencia de anuncio sea sensible.

En este paso del proceso, tenemos que seleccionar un umbral de puntuación de similitud relativamente alto que nos permita crear un subconjunto de preferencias publicitarias probablemente sensibles que se puedan etiquetar manualmente con un esfuerzo manual razonable.

La Figura 4 muestra la CDF para la puntuación de similitud semántica de las 126K preferencias publicitarias. La curva es plana cerca de 0 y 1, con un fuerte aumento entre los valores de similitud 0.25 y 0.6. Este fuerte aumento implica que establecer nuestro umbral en valores por debajo de 0.6 daría lugar a un rápido crecimiento del número de preferencias publicitarias que se etiquetarán manualmente. Por lo tanto, establecemos el umbral de similitud semántica en 0.6 porque corresponde a una puntuación de similitud relativamente alta. El subconjunto filtrado automáticamente resultante incluye 4452 preferencias publicitarias (3.5% de 126K), que es un número razonable para ser etiquetado manualmente.

Tenga en cuenta que la CDF tiene dos saltos con puntuaciones de similitud iguales a 0.5 y 0.58. El primero está vinculado a la categoría de desambiguación *"Local Business"*, mientras que el segundo se refiere a la categoría de desambiguación *"Public Figure"*. En general, no esperamos encontrar un número significativo de preferencias publicitarias potencialmente sensibles dentro de estas categorías de desambiguación. Por lo tanto, esta observación refuerza nuestra selección de umbral de similitud semántica en 0.6.

5.1.2 Clasificación manual de preferencias publicitarias potencialmente sensibles

Reclutamos doce evaluadores. Todos ellos son investigadores (profesores y estudiantes de doctorado) con algún conocimiento en el área de privacidad. Cada evaluador clasificó manualmente una muestra aleatoria (entre 1000 y 4452 elementos) de las 4452 preferencias publicitarias incluidas en el subconjunto filtrado automáticamente descrito anteriormente. Les pedimos que clasifiquen cada preferencia de anuncio en una de las cinco categorías consideradas sensibles (Política, Salud, Etnia, Religión, Sexualidad), en la categoría "Otro" (si no corresponde a ninguna de las categorías sensibles), o en la categoría "Desconocido" (si el evaluador no conoce el significado de la preferencia del anuncio). Para llevar a cabo el etiquetado manual, los investigadores recibieron toda la información contextual que Facebook ofrece por preferencia de anuncio: nombre, categoría de desambiguación y la categoría (si están disponibles).[12]

Cada preferencia de anuncio fue clasificada manualmente por cinco miembros. Usamos la votación por mayoría [20] para clasificar cada preferencia de anuncio como sensible o no sensible. Es decir, etiquetamos una preferencia de anuncio como sensible si al menos tres votantes (es decir, la mayoría) lo clasifican en una de las cinco categorías sensibles y, de lo contrario, no es sensible.

[12]Las instrucciones proporcionadas a los panelistas fueron: *"Asigne solo una categoría por preferencia de anuncio. Si cree que más de una categoría se aplica a una preferencia de anuncio, use solo la que considere más relevante. Si ninguna de las categorías coincide con la preferencia del anuncio, clasifíquela como 'Otra'. En caso de que no sepa el significado de la preferencia de un anuncio, lea la categoría de desambiguación y el topic que puede ayudarlo. Si después de leerlos sigue sin poder clasificar la preferencia de anuncio, use 'No se sabe' para clasificarla."*

votos	0	1	2	3	4	5
#preferencias	1054	767	539	422	449	1221

Table 1: Número de preferencias de anuncio que recibieron 0, 1, 2, 3, 4 o 5 votos clasificándolas dentro de una categoría sensible.

La tabla 1 muestra la cantidad de preferencias publicitarias que recibieron 0, 1, 2, 3, 4 y 5 votos clasificándolos en una categoría sensible. 2092 de las 4452 preferencias publicitarias están etiquetadas como sensibles, es decir, se han clasificado en una categoría sensible por al menos 3 votantes. Esto representa el 1.66% de las 126K preferencias publicitarias de nuestro conjunto de datos.

Una preferencia de anuncio clasificada como sensible puede haber sido asignada a diferentes categorías sensibles (por ejemplo, política y religión) por diferentes votantes. Hemos evaluado el acuerdo entre los votantes en las categorías sensibles asignadas a las preferencias publicitarias etiquetadas como sensibles mediante el test de Fleiss 'Kappa [10][11]. El coeficiente de Fleiss 'Kappa obtenido es de 0,94. Esto indica un acuerdo casi perfecto entre los votos de los evaluadores que vincula una preferencia de anuncios a una categoría sensible [16]. Por lo tanto, concluimos que (casi) todas las preferencias publicitarias clasificados como sensibles corresponden a una categoría sensible única entre las 5 consideradas.

Las 2092 preferencias publicitarias etiquetadas como sensibles se distribuyen de la siguiente manera en las cinco categorías sensibles: 58.3% están relacionadas con la política, 20.8% a la religión, 18.2% a la salud, 1.5% a la sexualidad, 1.1% a la etnia y solo el 0,2% presenta discrepancia entre los votos. Se puede acceder a la lista completa de las preferencias publicitarias clasificados como sensibles a través del sitio FDVT.[13] Nos referimos a este subconjunto de 2092 preferencias publicitarias como *subconjunto sospechosamente sensible*.

5.2 Obtención desde el Administrador de Anuncios de Facebook del número de usuarios etiquetados con preferencias publicitarias potencialmente sensibles

Aprovechamos la API Ads Manager de FB para obtener el número de usuarios de FB en cada país de la UE a los que se les ha asignado cada una de las 2092 preferencias publicitarias potencialmente sensibles del subconjunto sensible sospechoso. Recopilamos esta información en enero de 2018. A continuación, clasificamos estas preferencias publicitarias de las más populares a las menos populares de cada país. Esto nos permite calcular el número de usuarios de FB asignados con al menos una del Top N (con N entre 1 y 2092) preferencias publicitarias potencialmente sensibles. Para obtener esta información, utilizamos la operación OR disponible en la API de FB para crear audiencias. Esta función nos permite obtener cuántos usuarios de un país determinado están interesados en *preferencia de anuncio 1* O *preferencia de anuncio 2* O *preferencia de anuncio 3 ...* O *preferencia de anuncio N*. Un ejemplo de esto para N = 3 podría ser *"cuántas personas en Francia están interesadas en el comunismo O el Islam O el Veganismo"*.

Aunque el número de usuarios es una métrica relevante, no ofrece un resultado comparativo justo para evaluar la importancia del problema en todos los países porque podemos encontrar países de la UE con decenas de millones de usuarios (por ejemplo, Francia, Alemania, Italia, etc.) y algunos otros con menos de un millón (por ejemplo, Malta, Luxemburgo, etc.). Por lo tanto, utilizamos la proproporción de usuarios en cada país a los que se han asignado preferencias publicitarias potencialmente sensibles como la métrica para analizar los resultados. Más allá de los usuarios de FB, también estamos interesados en cuantificar la proporción de ciudadanos con preferencias publicitarias sensibles en cada país de la UE. Hemos definido dos métricas utilizadas en el resto del artículo:

- ***FFB (C, N)***: este es el porcentaje de usuarios de FB en el país C a los que se ha asignado al menos una de las Top N preferencias publicitarias del subconjunto sensible sospechoso. Tomamos nota de que C también puede referirse a los 28 países de la UE cuando deseamos analizar los resultados para toda la UE. Se calcula como la proproporción entre el número de usuarios de FB a los que se ha asignado al menos una de las N preferencias publicitarias más populares y el número total de usuarios de FB en el país C, que se puede obtener del Administrador de Anuncios de FB.

- ***FC (C, N)***: este es el porcentaje de ciudadanos en el país C (o en todos los países de la UE) a los que se les ha asignado al menos una de las Top N preferencias publicitarias potencialmente sensibles. Se calcula como la proproporción entre el número de ciudadanos a los que se ha asignado al menos una de las N preferencias publicitarias más populares y la población total del país C. Utilizamos datos del Banco Mundial para obtener las poblaciones de los países de la UE.[14]

El criterio para seleccionar las preferencias publicitarias Top N de las 2092 preferencias publicitarias potencialmente sensibles identificadas es la popularidad. Esto significa que seleccionamos las N preferencias publicitarias asignadas a la mayoría de los usuarios de acuerdo con la API Ads Manager de FB. Tenga en cuenta que

[13] `https://fdvt.org/usenix2018/panelists.html`

[14] `https://data.worldbank.org`

razones de asignación	todas las preferencias	las potencialmente sensibles
debido a un "me gusta"	71.64%	81.36%
debido a un clic en un anuncio	21.51%	15.85%
FB sugiere que podría ser relevante	4.83%	2.45%
debido a la instalación de una app	1.78%	0.04%
debido a comentarios o reacciones	0.18%	0.26%
añadida por el usuario	0.04%	0.03%
inciertas o no obtenidas por el FDVT	0.01%	0.01%

Table 2: Frecuencia de las seis razones por las cuales las preferencias publicitarias son asignadas a usuarios del FDVT de la UE de acuerdo con las explicaciones proporcionadas por FB.

FFB (C, N) y FC (C, N) probablemente informarán un límite inferior con respecto al porcentaje total de usuarios y ciudadanos de FB en el país C etiquetados con preferencias publicitarias potencialmente sensibles por dos razones. En primer lugar, estas métricas pueden utilizar a lo sumo N=2092 preferencias publicitarias potencialmente sensibles, lo que (suponiendo que nuestros votantes son precisos) es muy probablemente un subconjunto de todas las preferencias publicitarias sensibles disponibles en FB. En segundo lugar, la API de FB solo permite crear audiencias con un máximo de N=1000 intereses (es decir, preferencias publicitarias). Más allá de N=1000 intereses, la API proporciona un número fijo de usuarios de FB independientemente de la audiencia definida. Este número fijo es 2.1B que, según nuestro conocimiento, hace referencia al número total de usuarios de FB incluidos en el Administrador de Anuncios. Por lo tanto, en la práctica, el valor máximo de N que podemos usar en FFB y FC es 1000.

6 Cuantificando la exposición de usuarios de la UE a preferencias publicitarias potencialmente sensibles

En esta sección, primero analizamos la exposición de los usuarios del FDVT a las 2092 preferencias publicitarias potencialmente sensibles incluidas en el subconjunto de sospechosas sensibles. Posteriormente, utilizamos las métricas FFB y FC para analizar la exposición de los usuarios y ciudadanos de FB de la UE a esas preferencias publicitarias. Finalmente, realizamos un análisis demográfico para comprender si los usuarios de ciertos grupos de género o edad están más expuestos a preferencias publicitarias sensibles.

6.1 Usuarios del FDVT

4121 (90%) usuarios del FDVT están etiquetados con al menos una preferencia de anuncio sensible. En general, las 2092 preferencias publicitarias sensibles se han asignado más de 146K veces a los usuarios del FDVT. Si nos centramos solo en los usuarios de la UE, que son el enfoque de este documento, 2848 (90%) han sido etiquetados con preferencias publicitarias potencialmente sensibles. En general, se les ha asignado más de 100K intereses sensibles (1528 únicos). El número mediano (media) de preferencias publicitarias potencialmente sensibles asignadas a los usuarios del FDVT es 10 (16). Los percentiles 25 y 75 son 5 y 21, respectivamente.

Nuestro conjunto de datos del FDVT incluye la razón por la cual, según FB, cada preferencia de anuncio se ha asignado a un usuario. La Tabla 2 muestra la frecuencia de cada razón para todas las preferencias publicitarias y solo las potencialmente sensibles. Los resultados indican que la mayoría de las preferencias publicitarias sensibles se derivan de *"me gusta" de los usuarios* (81%) o *clics en los anuncios* (16%). Hay muy pocos casos (0.03%) en los que los usuarios incluyen de forma proactiva las preferencias publicitarias potencialmente sensibles en su lista de preferencias publicitarias utilizando la configuración que ofrece FB. Como recordatorio, según el RGPD de la UE, FB debe obtener un permiso explícito para procesar y explotar datos personales sensibles. Los *"me gusta"* de los usuarios y los clics en los anuncios no parecen cumplir este requisito.

6.2 Usuarios de FB y ciudadanos de la UE

La Figura 5 muestra el FFB(C, N) para valores de N que oscilan entre 1 y 1000. La figura informa los valores máximo, mínimo y promedio en los 28 países de la UE.[15] Observamos que incluso si consideramos un número bajo de preferencias publicitarias sensibles, la fracción de usuarios afectados es muy significativa. Por ejemplo, en promedio, el 60% de los usuarios de FB de los países de la UE están etiquetados con alguna de las 10 principales (es decir, las más populares) preferencias publicitarias potencialmente sensibles.

Además, observamos que el FFB es estable para valores de N que oscilan entre 500 y 1000. Observamos que hemos obtenido el mismo resultado estable para cada país de la UE. Esto indica que es probable que ya se haya etiquetado a cualquier usuario con preferencias publicitarias potencialmente sensibles fuera de las top500.[16] Suponemos que este comportamiento asintótico puede indicar que el límite inferior representado por FFB(C,N=500) está cerca de la fracción real de usuarios de FB etiquetados con preferencias publicitarias sensibles.

[15] La media en los países de la UE se ha calculado sumando el promedio de cada país de la UE y dividiéndolo por 28, ya que la preferencia Top N para cada país cambia de un país a otro.

[16] La lista de las 500 principales por país en `https://fdvt.org/usenix2018/top500.html`

país	C	FFB(C,500)	FC (C,500)	país	C	FFB(C,500)	FC (C,500)
Austria	AT	75.00	37.73	Irlanda	IE	80.65	52.38
Bélgica	BE	70.27	45.82	Italia	IT	79.41	44.55
Bulgaria	BG	72.97	37.88	Letonia	LV	72.53	33.67
Croacia	HR	80.00	38.36	Lituania	LT	75.00	41.78
Chipre	CY	79.17	64.95	Luxemburgo	LU	72.22	44.60
República Checa	CZ	71.70	35.98	Malta	MT	80.56	66.37
Dinamarca	DK	77.50	54.09	Holanda	NL	74.55	48.18
Estonia	EE	66.67	36.46	Polonia	PL	75.00	31.62
Finlandia	FI	70.97	40.04	Portugal	PT	81.54	51.33
Francia	FR	65.79	37.37	Rumanía	RO	75.76	38.06
Alemania	DE	67.57	30.24	España	ES	74.07	43.06
Reino Unido	GB	75.00	50.28	Eslovaquia	SK	70.37	35.00
Grecia	GR	77.19	40.94	Eslovenia	SI	78.00	37.78
Hungría	HU	75.44	43.80	Suecia	SE	73.97	54.53
				Unión Europea	UE	73.25	40.63

Table 3: Porcentaje de usuarios de FB (FFB) y ciudadanos (FC) de la UE por país que han sido etiquetados con alguna de las preferencias potencialmente sensibles del Top 500. La última fila reporta el valor agregado para los 28 países juntos de la UE.

La Tabla 3 muestra FFB(C,N=500) y FC(C,N=500) para cada país de la UE. La última fila de la tabla muestra los resultados promedio de los 28 países de la UE juntos (UE28).

Observamos que el 73% de los usuarios de FB de la UE, que corresponde al 40% de los ciudadanos de la UE, están etiquetados con alguna de las 500 principales preferencias publicitarias potencialmente sensibles en nuestro conjunto de datos. Si nos centramos en países individuales, FC(C,N=500) revela que en 7 de ellos, más de la mitad de sus ciudadanos están etiquetados con al menos una de las 500 principales preferencias potencialmente sensibles: Malta (66,37%), Chipre (64,95%), Suecia (54,53%), Dinamarca (54,09%), Irlanda (52,38%), Portugal (51,33%) y Gran Bretaña (50,28%). En contraste, los 5 países menos afectados son: Alemania (30,24%), Polonia (31,62%), Letonia (33,67%), Eslovaquia (35%) y República Checa (35,98%). Además, el FFB(C,N=500) oscila entre el 65% para Francia y el 81% para Portugal. Esto significa que aproximadamente 2/3 o más de los usuarios de FB en cualquier país de la UE están etiquetados con alguna de las 500 preferencias publicitarias potencialmente sensibles más populares.

Estos resultados sugieren que una gran parte de la población de la UE puede ser objeto de campañas publicitarias basadas en datos personales potencialmente delicados.

6.3 Preferencias publicitarias sensibles verificadas por un experto

Para confirmar que nuestro conjunto de preferencias publicitarias potencialmente sensibles contiene las que probablemente sean relevantes bajo el RGPD, examinamos un subconjunto de 20 preferencias publicitarias que todos los evaluadores clasificaron como confidenciales. Un experto de la AEPD española revisó y confirmó la sensibilidad de cada una de las 20 preferencias publicitarias en ese subconjunto de acuerdo con el RGPD. Destacamos que este subconjunto no es necesariamente representativo de todas las preferencias publicitarias potencialmente sensibles (o preferencias que los ciudadanos de la UE pueden considerar delicadas), pero representa un subconjunto validado por un experto que utilizamos para un análisis más detallado.

Las Tablas 4 y 5 muestran el porcentaje de usuarios FB (FFB) y ciudadanos (FC) por país de la UE etiquetados con cada una de las 20 preferencias publicitarias sensibles verificadas. Tenga en cuenta que la última fila presenta los resultados agregados para las 20 preferencias en cada país, y la última columna presenta los resultados agregados para los 28 países de la UE juntos.

Observamos que el 42.9% de los usuarios de FB de la UE, que corresponde al 23.5% de los ciudadanos de la UE, están etiquetados con al menos una de las preferencias publicitarias sensibles verificadas. Por lo tanto, alrededor de una cuarta parte de la población de la UE ha sido etiquetada en FB con al menos una de las preferencias publicitarias sensibles verificadas. Si analizamos los resultados por país, observamos que la fracción de la población afectada oscila entre el 15% en Estonia (EE), Letonia (LV) y Polonia (PL) y el 38% en Malta (MT). Estos hallazgos sugieren que FB pudo haber usado datos delicados según el RGPD para un gran porcentaje de ciudadanos de la UE en el periodo anterior a la fecha en que se hizo ejecutable el RGPD.

6.4 Análisis demográfico de edad y género

Analizamos la asociación de diferentes grupos demográficos (según el género y la edad) con preferencias publicitarias potencialmente sensibles. El análisis de género considera dos grupos, hombres contra mujeres, mientras que el análisis de edad considera cuatro grupos de edad siguiendo la división propuesta por Erikson et al. [7]: 13-19 (Adolescencia), 20-39 (Edad adulta temprana), 40-64 (Edad adulta) y 65+ (Madurez). Para cada grupo, calculamos FFB(C=UE28,N=500) a partir del subconjunto de 2092 preferencias publicitarias sensibles sospechosas y FFB(C=UE28,N=20) utilizando preferencias publicitarias sensibles verificadas exclusivamente por el experto de la AEPD. Las Figuras 6 y 7 muestran los resultados para los grupos de edad y género, respectivamente.

El grupo Edad Adulta Temprana es claramente el grupo de edad más expuesto a las preferencias publicitarias sensibles sospechosas (o las 20 verificadas). 61% (45%) de los usuarios de este grupo han sido etiquetados con alguna de las preferencias publicitarias sensibles de las 500 principales sospechosas (20 verificadas). Después, encontramos los grupos de Adolescencia, Edad Adulta y Madurez con 55% (42%), 40% (32%) y 39% (28%) de sus usuarios etiquetados con algunas de las Top

nombre	AT	BE	BG	HR	CY	CZ	DK	EE	FI	FR	DE	GR	HU	IE	IT	LV	LT	LU	MT	NL	PL	PT	RO	SK	SI	ES	SE	GB	UE28
COMUNISMO	0.48	0.61	1.35	1.30	1.67	3.21	0.38	0.61	0.52	2.29	0.43	0.81	0.74	0.52	1.15	0.56	0.94	0.64	0.39	0.24	2.19	0.94	1.90	1.74	1.70	0.56	0.30	0.41	0.93
ISLAM	8.18	7.16	4.59	5.50	13.54	4.91	6.75	2.22	4.19	7.89	7.57	4.21	2.28	4.19	4.12	2.75	2.38	5.00	6.67	5.36	2.44	3.69	3.50	3.11	6.50	4.07	6.58	6.82	5.71
CORÁN	3.41	3.38	1.08	1.00	4.48	0.45	1.90	0.65	1.16	3.95	3.24	1.18	0.74	1.35	1.71	1.01	0.51	1.83	1.86	2.45	0.45	0.62	0.77	0.56	2.00	0.96	2.74	3.64	2.46
PREVENCIÓN DEL SUICIDIO	0.14	0.15	0.20	0.32	0.21	0.12	0.12	0.10	0.09	0.16	0.14	0.23	0.12	1.10	0.28	0.13	0.15	0.28	0.27	0.15	0.14	0.22	0.13	0.44	0.26	0.44	0.15	0.27	0.28
SOCIALISMO	1.00	0.78	0.57	0.48	1.15	2.45	3.00	0.76	0.48	0.47	0.43	0.91	1.93	1.10	3.53	0.34	0.94	2.78	1.08	0.28	0.50	2.15	0.35	2.33	0.82	1.48	1.37	0.93	1.21
JUDAÍSMO	2.50	1.16	0.86	0.70	2.29	0.72	2.17	1.01	0.61	1.26	1.38	1.30	1.16	1.26	2.29	1.76	1.81	1.19	3.06	1.00	1.19	1.69	1.40	0.93	0.74	1.15	0.64	0.95	1.32
HOMOSEXUALIDAD	6.14	5.54	2.97	6.50	4.38	5.47	5.00	3.89	5.16	7.37	5.68	5.09	4.21	9.03	7.65	4.62	3.19	5.00	7.50	6.18	3.56	4.46	3.80	4.44	7.60	8.15	4.93	8.64	6.79
MEDICINA ALTERNATIVA	5.00	2.97	8.38	6.00	5.62	4.15	4.00	4.17	4.19	2.89	3.24	7.19	4.21	9.68	6.18	3.96	2.56	5.56	7.50	3.64	2.25	8.00	3.90	2.93	5.00	5.56	3.84	6.14	4.29
CRISTIANISMO	10.68	7.43	6.22	7.50	9.69	3.77	15.00	2.22	4.19	5.53	6.49	6.67	9.30	10.97	12.65	3.19	3.81	7.22	18.89	5.18	6.25	12.46	10.00	4.81	4.60	10.00	4.66	7.50	8.21
INMIGRACIÓN ILEGAL	0.17	0.07	0.10	0.02	0.07	0.68	0.05	0.01	0.07	0.05	0.06	0.26	0.26	0.06	0.08	0.02	0.06	0.01	0.08	0.02	0.02	0.02	0.02	0.11	0.36	0.14	0.33	0.05	0.09
ONCOLOGÍA	0.23	0.27	0.62	0.44	3.96	0.57	0.15	0.10	0.08	0.17	0.16	0.49	0.30	1.29	0.94	0.70	1.62	0.19	0.78	0.45	1.25	1.09	0.73	0.59	0.21	0.70	0.08	0.66	0.61
COMUNIDAD LGBT	6.36	6.62	5.14	6.50	6.56	6.04	6.50	5.14	6.45	7.11	5.95	5.79	4.39	11.94	8.53	5.27	5.88	6.67	9.44	6.36	5.88	7.85	6.30	4.81	6.00	7.04	6.44	11.14	8.21
IDENTIDAD DE GÉNERO	0.03	0.08	0.01	0.08	0.88	0.02	0.03	0.02	0.02	0.07	0.03	0.56	0.07	0.23	0.07	0.20	0.10	0.10	0.14	0.03	0.05	0.05	0.04	0.01	0.08	0.07	0.09	0.55	0.10
SALUD REPRODUCTIVA	0.01	0.07	0.20	0.40	0.02	0.14	0.05	0.02	0.06	0.01	0.01	0.04	0.10	0.71	0.04	0.07	0.05	0.01	0.24	0.03	0.01	0.04	0.01	0.03	0.00	0.03	0.05	0.13	0.07
BIBLIA	17.95	10.81	8.65	10.50	11.46	7.17	12.75	4.31	4.84	7.63	15.41	8.25	10.00	19.03	17.65	5.71	6.25	14.44	20.28	10.91	14.38	12.31	8.70	6.67	7.40	7.04	5.48	15.68	12.14
EMBARAZO	15.68	12.97	9.19	17.00	13.54	16.23	14.50	10.00	11.29	10.79	11.89	13.51	11.23	20.97	12.35	13.19	18.75	12.78	9.72	14.55	15.00	18.46	9.70	18.89	13.00	14.07	13.42	18.41	14.29
NACIONALISMO	0.86	0.78	1.65	1.85	2.19	2.45	1.00	0.58	0.45	1.08	1.00	1.74	2.11	2.00	1.32	2.42	0.94	2.19	2.78	0.70	3.00	1.69	2.50	1.37	0.61	1.11	0.99	0.91	1.39
VEGANISMO	14.55	10.27	7.30	10.50	10.21	9.25	12.75	9.86	15.16	8.68	11.35	9.82	9.82	14.84	13.53	9.23	8.12	13.06	13.33	10.91	8.12	11.23	6.70	8.52	14.00	10.37	16.44	13.64	11.43
BUDISMO	3.18	3.38	1.62	3.55	3.33	2.26	2.08	1.53	1.13	2.61	1.43	2.63	3.33	3.87	2.94	1.98	1.88	3.33	4.17	2.45	1.31	6.92	1.90	1.67	3.00	2.19	1.51	2.50	2.39
FEMINISMO	4.55	3.78	3.51	3.80	5.52	2.08	5.50	2.78	6.77	5.00	3.78	3.68	2.46	9.35	5.88	3.19	3.56	5.83	8.61	3.64	3.44	8.15	2.40	4.07	3.90	8.89	13.70	7.27	7.50
UNIÓN	45.45	39.19	32.43	41.50	45.83	37.74	45.00	27.78	35.48	34.21	40.54	36.84	36.84	51.61	44.12	32.97	36.25	41.67	47.22	40.00	36.88	44.62	34.34	35.56	39.00	40.74	41.10	47.73	42.86

Table 4: Porcentaje de usuarios de FB (FFB) para cada país de la UE que fueron etiquetados con cada una de las 20 preferencias de anuncio sensibles validadas por el experto de la AEPD. La última fila reporta el valor FFB agregado para las 20 preferencias de anuncio y para cada país de la UE. La última columna devuelve el valor agregado para los 28 países de la UE.

nombre	AT	BE	BG	HR	CY	CZ	DK	EE	FI	FR	DE	GR	HU	IE	IT	LV	LT	LU	MT	NL	PL	PT	RO	SK	SI	ES	SE	GB	UE28
COMUNISMO	0.24	0.40	0.70	0.62	1.37	1.61	0.26	0.33	0.29	1.30	0.19	0.43	0.43	0.34	0.64	0.26	0.52	0.39	0.32	0.15	0.92	0.59	0.96	0.87	0.82	0.32	0.22	0.27	0.51
ISLAM	4.12	4.67	2.39	2.64	11.11	2.46	4.71	1.22	2.37	4.48	3.39	2.23	1.32	2.72	2.31	1.28	1.32	3.09	5.49	3.47	1.03	2.32	1.78	1.55	3.15	2.37	4.85	4.57	3.13
CORÁN	1.71	2.20	0.56	0.48	3.67	0.23	1.33	0.36	0.66	2.24	1.45	0.62	0.43	0.88	0.96	0.47	0.28	1.13	1.53	1.59	0.19	0.39	0.39	0.28	0.97	0.56	2.02	2.44	1.35
PREVENCIÓN DEL SUICIDIO	0.07	0.10	0.10	0.15	0.17	0.06	0.08	0.05	0.05	0.09	0.06	0.12	0.07	0.71	0.16	0.06	0.08	0.17	0.22	0.09	0.06	0.14	0.07	0.22	0.13	0.26	0.11	0.18	0.15
SOCIALISMO	0.50	0.51	0.29	0.23	0.94	1.23	2.09	0.42	0.27	0.27	0.19	0.48	1.12	0.71	1.98	0.16	0.52	1.72	0.89	0.18	0.21	1.36	0.18	1.16	0.40	0.86	1.01	0.62	0.66
JUDAÍSMO	1.26	0.76	0.45	0.34	1.88	0.36	1.52	0.55	0.35	0.72	0.62	0.69	0.67	0.82	1.29	0.82	1.01	0.74	2.52	0.65	0.50	1.07	0.71	0.46	0.36	0.67	0.47	0.64	0.72
HOMOSEXUALIDAD	3.09	3.61	1.54	3.12	3.59	2.75	3.49	2.13	2.91	4.19	2.54	2.70	2.44	5.87	4.29	2.14	1.78	3.09	6.18	4.00	1.50	2.81	1.93	2.21	3.68	4.74	3.64	5.79	3.71
MEDICINA ALTERNATIVA	2.52	1.94	4.35	2.88	4.61	2.08	2.79	2.28	2.37	1.64	1.45	3.82	2.44	6.29	3.47	1.84	1.43	3.43	6.18	2.35	0.95	5.04	1.98	1.46	2.42	3.23	2.83	4.11	2.34
CRISTIANISMO	5.37	4.85	3.23	3.60	7.95	1.89	10.47	1.22	2.37	3.14	2.90	3.54	5.40	7.12	7.10	1.48	2.12	4.46	15.56	3.35	2.64	7.85	5.07	2.39	2.23	5.81	3.43	5.03	4.49
INMIGRACIÓN ILEGAL	0.09	0.04	0.05	0.01	0.06	0.34	0.03	0.00	0.04	0.03	0.03	0.14	0.15	0.04	0.04	0.01	0.03	0.01	0.07	0.01	0.01	0.01	0.01	0.05	0.17	0.08	0.24	0.04	0.05
ONCOLOGÍA	0.11	0.18	0.32	0.21	3.25	0.28	0.10	0.06	0.05	0.10	0.07	0.26	0.17	0.84	0.53	0.33	0.91	0.12	0.64	0.29	0.53	0.69	0.37	0.29	0.10	0.41	0.06	0.44	0.33
COMUNIDAD LGBT	3.20	4.32	2.67	3.12	5.38	3.03	4.54	2.81	3.64	4.04	2.66	3.07	2.55	7.75	4.79	2.45	3.27	4.12	7.78	4.11	2.48	4.94	3.20	2.39	2.91	4.09	4.75	7.47	4.49
IDENTIDAD DE GÉNERO	0.01	0.05	0.01	0.04	0.72	0.01	0.02	0.01	0.01	0.04	0.01	0.30	0.04	0.15	0.04	0.09	0.06	0.06	0.12	0.02	0.02	0.03	0.02	0.00	0.04	0.04	0.06	0.37	0.05
SALUD REPRODUCTIVA	0.00	0.05	0.11	0.19	0.02	0.07	0.04	0.01	0.04	0.01	0.00	0.02	0.06	0.46	0.02	0.03	0.03	0.01	0.19	0.02	0.01	0.02	0.01	0.01	0.00	0.02	0.03	0.09	0.04
BIBLIA	9.03	7.05	4.49	5.04	9.40	3.60	8.90	2.35	2.73	4.34	6.90	4.37	5.81	12.36	9.90	2.65	3.48	8.92	16.71	7.05	6.06	7.75	4.42	3.32	3.58	4.09	4.04	10.51	6.64
EMBARAZO	7.89	8.46	4.77	8.15	11.11	8.14	10.12	5.47	6.37	6.13	5.32	7.16	6.52	13.62	6.93	6.12	10.44	7.89	8.01	9.40	6.32	11.62	4.92	9.39	6.30	8.18	9.90	12.34	7.82
NACIONALISMO	0.43	0.51	0.86	0.89	1.79	1.23	0.70	0.32	0.25	0.61	0.45	0.92	1.22	1.30	0.74	1.12	0.52	1.36	2.29	0.45	1.26	1.07	1.27	0.68	0.30	0.65	0.73	0.61	0.76
VEGANISMO	7.32	6.70	3.79	5.04	8.38	4.64	8.90	5.39	8.55	4.93	5.08	5.21	5.70	9.64	7.59	4.28	4.53	8.06	10.99	7.05	3.43	7.07	3.40	4.24	6.78	6.03	12.12	9.14	6.25
BUDISMO	1.60	2.20	0.84	1.70	2.73	1.14	1.45	0.84	0.64	1.48	0.64	1.40	1.94	2.51	1.65	0.92	1.04	2.06	3.43	1.59	0.55	4.36	0.96	0.83	1.45	1.27	1.11	1.68	1.31
FEMINISMO	2.29	2.47	1.82	1.82	4.53	1.04	3.84	1.52	3.82	2.84	1.69	1.95	1.43	6.08	3.30	1.48	1.98	3.60	7.09	2.35	1.45	5.13	1.22	2.03	1.89	5.17	10.10	4.88	4.10
UNIÓN	22.86	25.55	16.84	19.90	37.60	18.94	31.41	15.19	20.02	19.43	18.14	19.54	21.39	33.52	24.75	15.30	20.19	25.73	38.91	25.85	15.55	28.09	17.25	17.68	18.89	23.68	30.29	31.99	23.45

Table 5: Porcentaje de ciudadanos (FC) para cada país de la UE que fueron etiquetados con cada una de las 20 preferencias de anuncio sensibles validadas por el experto de la AEPD. La última fila reporta el valor FC agregado para las 20 preferencias de anuncio y para cada país de la UE. La última columna devuelve el valor agregado para los 28 países de la UE.

500 preferencias publicitarias sensibles (20 verificadas), respectivamente. Aunque la diferencia en la exposición a las preferencias publicitarias sensibles es sustancial entre los grupos, todos ellos presentan una exposición considerablemente alta. En particular, más de una cuarta parte de los usuarios de cada grupo están expuestos a preferencias publicitarias sensibles verificadas por el experto.

El análisis de género muestra que el 78% (49%) de las mujeres están expuestas a las preferencias publicitarias del Top 500 (20 verificadas). La exposición es notablemente más pequeña para los hombres, donde la fracción de usuarios etiquetados con alguna de las preferencias publicitarias sensibles Top 500 (20 verificadas) se reduce en 10 (18) puntos porcentuales a 68% (31%). Este resultado sugiere la existencia de un sesgo de género, que a pesar de su interés está fuera del alcance de este artículo.

7 Explotación comercial de las preferencias de anuncio sensibles a través de campañas reales en FB

Nuestro análisis muestra que Facebook etiquetó a una parte significativa de los ciudadanos de la UE usando datos personales potencialmente sensibles. En esta sección, demostramos que FB permitió que fueran dirigidos anuncios a usuarios con preferencias publicitarias sensibles asignadas. Entre el 6 de octubre y el 15 de octubre de 2017, realizamos tres campañas publicitarias de FB que utilizaban preferencias sensibles verificadas como: *"creencias religiosas"* (dirigidas a usuarios interesados en el Islam O el judaísmo O el cristianismo O el budismo) *"opiniones políticas"* (dirigidas a usuarios interesados en comunismo O anarquismo O feminismo radical O socialismo) y *"orientación sexual"* (dirigidas a usuarios interesados en transexualidad O homosexualidad).[17] Las 3 campañas se enfocaron en cuatro países de la UE: Alemania, España, Francia e Italia.

En general, con un presupuesto de 35€ pudimos llegar a 26458 usuarios etiquetados con algunas de las preferencias publicitarias sensibles anteriores. Se cargó a nuestra tarjeta de crédito y recibimos las facturas y los informes de resumen asociados con nuestras campañas (consulte la Figura 8). Este experimento proporciona

[17] *"Anarquismo"* y *"Transexualismo"* no fueron verificados explícitamente por el experto, aunque reflejan fielmente las preferencias publicitarias verificadas.

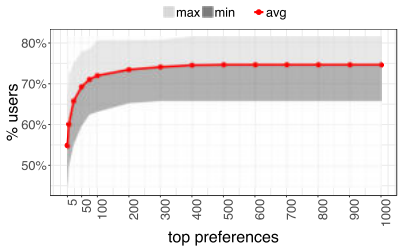


Figure 5: FFB (C,N) para valores de N entre 1 y 1000. La figura muestra el mínimo, media y máximo FFB para los 28 países de la UE.

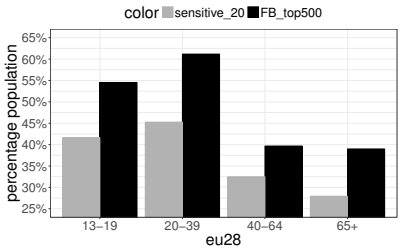


Figure 6: Porcentaje de usuarios de FB de la UE etiquetados con al menos una de las Top 500 (negro) y 20 verificadas (gris) preferencias de anuncio en los siguientes grupos de edad: 13-19, 20-39, 40-64, 65+.

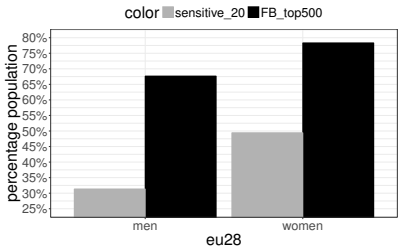


Figure 7: Porcentaje de usuarios de FB de la UE etiquetados con al menos una de las Top 500 (negro) y 20 verificadas (gris) preferencias de anuncio en los siguientes grupos de género: Hombre, Mujer.

Ad Set Name	Reach	Impressions	Amount Spent	Location (Ad Set Settings)
Religion	7,630	7,985	€5.00 of €5.00	IT, ES, FR and DE
Political	11,025	16,537	€10.00 of €10.00	IT, ES, FR and DE
Sexuality	7,314	7,367	€20.00 of €20.00	IT, ES, FR and DE
Results from 3 ad sets	26,458 People	31,889 Total	€35.00 Total Spent	

Figure 8: Resultados reportados por FB para las 3 campañas publicitarias que ejecutamos dirigiéndonos a usuarios etiquetados con preferencias de publicidad sensibles.

evidencia sustancial de que FB generó (antes del 25 de mayo) ingresos de la explotación comercial de datos personales sensibles verificados de acuerdo con la definición de RGPD de *datos protegidos*.

Según nuestro conocimiento, las campañas realizadas cumplían con los Términos de uso de Facebook presentados en la Sección 3.3 y el Apéndice E.

8 Riesgos éticos y de privacidad asociados con la explotación de datos personales sensibles

La posibilidad de llegar a usuarios etiquetados con datos personales potencialmente sensibles permite el uso de campañas publicitarias de FB para atacar a grupos específicos de personas basándose en datos personales delicados (raza, orientación sexual, creencias religiosas, etc.). A continuación, ilustramos dos ejemplos específicos de posibles ataques.

Campañas de odio: Un atacante podría crear campañas de incitación al odio usando preferencias publicitarias sensibles representativas de un grupo social sensible específico dentro de su público objetivo. Por ejemplo, una organización neonazi podría crear campañas publicitarias con mensajes ofensivos dirigidos a personas interesadas en *Judaísmo* u *Homosexualidad*. Las campañas de discurso de odio pueden llegar a miles de usuarios a un costo muy bajo (por ejemplo, nosotros llegamos a más de 26K usuarios de FB gastando solo 35€ en campañas de anuncios de FB).

Nuestro experimento en la Sección 7 muestra que tales campañas de odio pueden llegar a miles de usuarios a un coste muy bajo en el orden de decenas de euros.

Ataque de identificación: Un atacante puede usar FB para identificar a los ciudadanos que pertenecen a un grupo social sensible definido por su creencia religiosa, orientación sexual, preferencia política, etc. Para este fin, un atacante solo necesita replicar un ataque phishing [14]. El atacante configuraría una campaña dirigida a un público sensible (por ejemplo, personas interesadas en *homosexualidad*) utilizando un anuncio sofisticado que sirve como cebo para atraer a los usuarios objetivo a la página web del atacante (por ejemplo, el anuncio promete que el usuario ganará una iPhone X si hace clic en el anuncio). Si el usuario hace clic en el anuncio, será redirigido a la página web del atacante. Una vez allí, el atacante puede utilizar diferentes técnicas explotadas en los ataques de phishing [14] para persuadir al usuario de que proporcione algunos datos personales que revelarían su identidad. Por ejemplo, en el ejemplo del iPhone X, la página de destino puede mostrar un mensaje felicitando al usuario por haber ganado el teléfono solicitando que proporcione datos personales (nombre, dirección, número de teléfono, etc.) para fines de envío.

Un estudio reciente [13] realizó experimentos implementando ataques de phishing basados en correos electrónicos en los que el 9% de los usuarios publicaron sus credenciales (nombre de usuario y contraseña) en el sitio de phishing (es decir, la página de inicio del atacante). Usando como referencia esta tasa de éxito para los ataques de phishing y los resultados de las campañas

publicitarias descritas en la Sección 7, podemos hacer una estimación del coste de la identificación de usuarios etiquetados con preferencias publicitarias sensibles verificadas. Gastamos 35€ en nuestras campañas publicitarias para llegar a 26K usuarios, de los cuales 2,34K (según la tasa de éxito de referencia del 9% de la literatura) pueden proporcionar información personal en la página web del atacante que podría revelar su identidad. En base a esto, la identificación de un miembro arbitrario del grupo puede ser tan barata como 0.015€. Incluso si consideramos una tasa de éxito dos órdenes de magnitud menor (0.09%), el coste sería 1.5€ por usuario.

El coste estimado para revelar la identidad de los usuarios en base a datos personales potencialmente sensibles es bastante bajo, considerando los graves riesgos de privacidad a los que los usuarios se pueden enfrentar. Por ejemplo, (i) en países donde la homosexualidad se considera ilegal o los gobiernos inmorales u otras organizaciones podrían obtener la identidad de personas que probablemente son homosexuales (por ejemplo, interesados en *homosexualidad, LGBT*, etc.); (ii) las organizaciones neonazis podrían identificar personas en regiones específicas (apuntando a una ciudad o incluso a un código postal) que probablemente sean judíos (por ejemplo, interesados en *judaísmo, Shabbat*, etc.); (iii) las compañías de seguros de salud podrían tratar de identificar a las personas que pueden tener hábitos no rentables (por ejemplo, interesados en *tabaco, comida rápida*, etc.) o problemas de salud (por ejemplo, *intolerancia alimenticia*) para rechazarlos como clientes o cobrarles más por el seguro de salud. Los usuarios pueden enfrentar las consecuencias negativas de tales ataques tipo phishing, incluso si FB los ha etiquetado erróneamente con alguna preferencia de anuncio sensible. En resumen, aunque Facebook no permite que terceros identifiquen usuarios individuales directamente, las preferencias publicitarias pueden usarse como un proxy muy poderoso para realizar ataques de identificación[18] basados en datos personales potencialmente sensibles a un bajo coste. Tenga en cuenta que simplemente hemos descrito este ataque de phishing basado en anuncios, pero no lo hemos implementado debido a las implicaciones éticas.

9 Extensión del FDVT para informar a los usuarios de sus preferencias de anuncio potencialmente sensibles

Los resultados de las secciones anteriores motivan la necesidad de soluciones que informen a los usuarios sobre el uso de datos personales sensibles con fines publicitarios. Con este fin, hemos ampliado la extensión de navegador FDVT para informar a los usuarios sobre las preferencias publicitarias potencialmente sensibles que FB les ha asignado: (i) hemos creado un clasificador para etiquetar automáticamente las preferencias publicitarias asignadas a los usuarios de FDVT como confidenciales o no sensible; (ii) hemos modificado el back-end FDVT y el front-end para incorporar esta nueva característica.

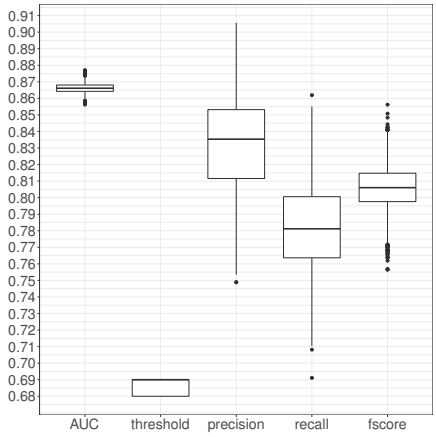


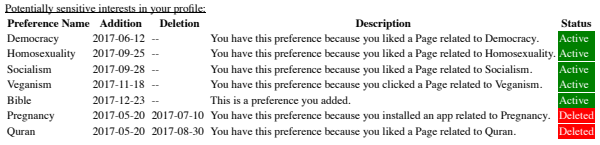
Figure 9: AUC, precisión, recall and F-score para el umbral óptimo que clasifica automáticamente una preferencia de anuncio como sensible o no sensible. La figura muestra los resultados obtenidos a partir de 5000 iteraciones para diferentes subconjuntos de datos de entrenamiento y validación seleccionados aleatoriamente.

9.1 Clasificador binario automático de preferencias de anuncio sensibles

Nos basamos en la metodología descrita en la Sección 5 para calcular la similitud semántica entre las preferencias publicitarias y las categorías de datos personales sensibles (es decir, política, religión, salud, etnia y orientación sexual). Recuerde que a cada preferencia de anuncio se le asigna una puntuación de similitud semántica que oscila entre 0 (el más bajo) y 1 (el más alto). Para crear un clasificador binario automático, debemos definir un umbral para que las preferencias publicitarias sobre el cual se clasifiquen como sensibles (o no sensibles).

Para establecer este umbral, usamos el conjunto de datos filtrados automáticamente de la Sección 5.1.2. Incluye 4452 preferencias publicitarias, donde 2092 se clasificaron como sensibles a partir de los votos de 12 evaluadores (es decir, subconjunto con sospecha de sensibilidad). Seguimos un modelo estándar de entrenamiento-prueba. Dividimos al azar nuestro conjunto de datos en subconjuntos de entrenamiento y validación que incluyen 80% y 20% de las muestras, respec-

[18]El ataque descrito puede implementarse en cualquier plataforma publicitaria que permita a los anunciantes dirigirse a los usuarios en función de datos personales delicados.



Potentially sensitive interests in your profile:

Preference Name	Addition	Deletion	Description	Status
Democracy	2017-06-12	--	You have this preference because you liked a Page related to Democracy.	Active
Homosexuality	2017-09-25	--	You have this preference because you liked a Page related to Homosexuality.	Active
Socialism	2017-09-28	--	You have this preference because you liked a Page related to Socialism.	Active
Veganism	2017-11-18	--	You have this preference because you clicked a Page related to Veganism.	Active
Bible	2017-12-23	--	This is a preference you added.	Active
Pregnancy	2017-05-20	2017-07-10	You have this preference because you installed an app related to Pregnancy.	Deleted
Quran	2017-05-20	2017-08-30	You have this preference because you liked a Page related to Quran.	Deleted

Figure 10: Sitio web que muestra preferencias de anuncio sensibles.

tivamente. El subconjunto de entrenamiento se utiliza para encontrar el umbral óptimo. A su vez, utilizamos el subconjunto de validación para evaluar el rendimiento del umbral seleccionado. El umbral óptimo se selecciona como el que maximiza el F-score para el subconjunto de entrenamiento [24]. Además, validamos el rendimiento del umbral seleccionado calculando la precisión, el recall y el F-score en el subconjunto de validación. Realizamos 5000 iteraciones de este proceso, cada una utilizando diferentes subconjuntos de prueba y validación elegidos al azar, para demostrar la robustez del clasificador binario propuesto.

La Figura 9 presenta diagramas de caja que muestran el AUC, la precisión, el recall y el F-score para el umbral óptimo en las 5000 iteraciones. El umbral óptimo permanece bastante estable entre 0,68 y 0,69. De manera similar, el AUC derivado de la curva ROC para nuestro clasificador binario presenta un resultado muy estable en torno a 0.86, que se asocia con un buen desempeño de acuerdo con las métricas de calidad estándar [9][28].

La precisión mediana de nuestro clasificador binario es 0.835 (min = 0.75, max = 0.90) y el recall mediano es de 0.78 (min = 0.70, max = 0.86).

Aunque el clasificador puede ser imperfecto, aún puede lograr el objetivo de aumentar la conciencia colectiva entre los usuarios de FB con respecto al uso potencial de datos personales sensibles con fines publicitarios.

9.2 Implementación

FDVT Backend: Calculamos la puntuación de similitud semántica para todas las preferencias publicitarias almacenados en nuestra base de datos. Para las preferencias publicitarias con una puntuación de similitud ≥ 0.69, las clasificamos como sensibles y las agregamos a una lista negra.[19] Cada vez que el usuario del FDVT inicia una sesión en FB. Recuperamos su conjunto actualizado de preferencias publicitarias y las comparamos con la lista negra para obtener una lista de preferencias publicitarias vinculadas a datos personales potencialmente sensibles. Almacenamos el historial de preferencias publicitarias potencialmente sensibles asignadas al usuario para notificarle aquellas preferencias que FB ha eliminado. Finalmente, cada vez que a un usuario se le asigna una nueva preferencia de anuncio que aún no está en nuestra base de datos, calculamos su puntuación de similitud semántica y la incluimos en la lista negra si la preferencia de anuncio se clasifica como sensible.

Interfaz de usuario FDVT: Hemos introducido un nuevo botón en la interfaz de extensión FDVT con la etiqueta *"Sensitive FB Preferences"*. Cuando un usuario hace clic en ese botón, mostramos una página web que enumera las preferencias publicitarias potencialmente sensibles incluidas en el conjunto de preferencias publicitarias del usuario. La Figura 10 muestra un ejemplo de esta página web. Proporcionamos la siguiente información para cada preferencia de anuncio potencialmente sensible: (i) Nombre de preferencia de anuncio, (ii) Fecha de adición, (iii) Fecha de eliminación (solo para preferencias publicitarias eliminados), (iv) Descripción, que indica la razón por la cual FB ha asignado esa preferencia de anuncio al usuario, y (v) Estado, ya sea activo (resaltado en verde) o eliminado (resaltado en rojo).

10 Estado del arte

Nos centramos en trabajos anteriores que abordan temas relacionados con datos personales sensibles en publicidad en línea, así como trabajos recientes que analizan temas de privacidad y discriminación relacionados con la publicidad y las preferencias publicitarias de FB.

Carrascosa et al. [4] propone una nueva metodología para cuantificar la proporción de anuncios dirigidos que reciben los usuarios de Internet mientras navegan por la web. Crean robots, denominados *personas*, con perfiles de intereses muy específicos (por ejemplo, personas interesadas en automóviles) y miden cuántos de los anuncios recibidos coinciden realmente con los intereses específicos de las personas analizadas. Crean personas basadas en datos personales sensibles (por ejemplo, salud) y demuestran que también están reciben anuncios relacionados con la información sensible utilizada para crear el perfil de la persona. Castellucia et al. [5] muestra que un atacante que obtiene acceso (por ejemplo, a través de una red WiFi pública) a los anuncios de Google recibidos por un usuario podría crear un perfil de interés que podría revelar hasta el 58% de los intereses reales del usuario. Los autores afirman que si algunos de los intereses revelados son sensibles, esto podría implicar graves riesgos de privacidad para los usuarios.

Venkatadri et al. [26] y Speicher et al. [25] exponen las vulnerabilidades de privacidad y discriminación relacionadas con la publicidad de FB. En [26], los autores demuestran cómo un atacante puede usar JavaScript de seguimiento de terceros de Facebook para recuperar datos personales (por ejemplo, números de teléfonos móviles) asociados con los usuarios que visitan el sitio

[19] El valor del umbral óptimo puede cambiar con el tiempo ya que se volverá a calcular periódicamente.

web del atacante. Además, en [25] demuestran que las preferencias publicitarias de FB sensibles se pueden usar para aplicar discriminación negativa en campañas publicitarias (por ejemplo, excluyendo personas según su raza). Los autores también muestran que algunas preferencias publicitarias que inicialmente pueden no ser sensibles podrían ser utilizadas para discriminar en campañas publicitarias (por ejemplo, excluyendo a las personas interesadas en *Blacknews.com* que son potencialmente personas negras).

Finalmente, Andreou et al. [3] analiza si los motivos que usa FB para explicar por qué un usuario al que se le muestra un anuncio está alineado con la audiencia real a la que se dirige el anunciante. Para hacer esto, analizan la explicación que incluye Facebook en cada anuncio entregado denominado *"¿Por qué veo este anuncio?"*, esta explicación describe el público objetivo asociado con el anuncio mostrado. Del análisis de 79 anuncios, concluyen que en muchos casos las explicaciones proporcionadas son incompletas y, a veces, engañosas. También realizan un análisis cualitativo relacionado con las preferencias publicitarias asignadas a los usuarios de FB basándose en un pequeño conjunto de datos que incluye 9K preferencias publicitarias distribuidas entre 35 usuarios. Llegan a la conclusión de que las razones por las que se asignan las preferencias publicitarias carecen de una explicación detallada.

En resumen, la literatura existente sugiere que el ecosistema de publicidad en línea (más allá de Facebook) explota información personal sensible con fines comerciales. Además, el trabajo anterior destaca varios problemas de privacidad, discriminación y transparencia asociados con las preferencias publicitarias de FB. Nuestro trabajo complementa este cuerpo de literatura que cuantifica el número de usuarios en FB que pueden estar expuestos a la explotación comercial de sus datos personales sensibles.

11 Consentimiento de los usuarios del FDVT

El comité de ética de la institución de los autores ha proporcionado la aprobación para llevar a cabo la implementación del FDVT y las actividades de investigación derivadas del mismo.

Para cumplir con los estándares éticos y legales más rigurosos, durante el proceso de instalación del FDVT, un usuario debe: (i) leer y aceptar los Términos de uso[20] y la Política de privacidad,[21] y (ii) otorgan permiso explícito para usar la información almacenada (de forma anónima) para fines de investigacion.

[20] `https://www.fdvt.org/terms_of_use/`

[21] `https://www.fdvt.org/privacy_agreement.html`

Finalmente, también vale la pena señalar que no recopilamos información (ni personal ni no personal) de los usuarios que hicieron clic en los anuncios que usamos en las campañas publicitarias de FB que se describen en la Sección 7.

12 Conclusiones

Nuestros hallazgos sugieren que Facebook explotó comercialmente datos personales potencialmente sensibles con fines publicitarios a través de las preferencias publicitarias que asigna a sus usuarios. Facebook ya ha sido multado en España por esta práctica. El RGPD entró en vigor el 25 de mayo de 2018. Hemos estudiado los datos personales potencialmente sensibles que FB asignó a los usuarios de la UE en el período anterior a esta fecha. Los resultados revelan que la proporción de usuarios de FB de la UE afectados es tan alta como el 73% (40% de ciudadanos de la UE). Hemos ilustrado cómo a los usuarios de FB a los que se les han asignado preferencias publicitarias sensibles podrían entrañar riesgos, como ataques dirigidos de bajo coste que buscan identificar a dichos usuarios. Los resultados de nuestro documento exigen una rápida reacción de Facebook para eliminar todas las preferencias publicitarias que se puedan usar para inferir la orientación política, la orientación sexual, las condiciones de salud, las creencias religiosas o el origen étnico de un usuario por dos razones: (i) esto puede evitar que Facebook incumpla el Artículo 9 del RGPD, y (ii) puede proteger a los usuarios de las amenazas que pueden surgir de la explotación de esta información delicada.

Agradecimientos

J.G. Cabañas agradece el financiamiento del Ministerio de Economía, Industria y Competitividad (España) a través del proyecto TEXEO (TEC2016-80339-R) y el Ministerio de Educación, Cultura y Deporte (España) a través de la Beca FPU (FPU16 / 05852). A. Cuevas agradece el financiamiento del Ministerio de Economía, Industria y Competitividad (España) y el Fondo Social Europeo (UE) a través de la Beca Ramón Y Cajal (RyC-2015-17732). R. Cuevas agradece la financiación del proyecto europeo H2020 SMOOTH (786741). También nos gustaría agradecer a los expertos legales que han proporcionado comentarios muy valiosos para este trabajo.

References

[1] Directive 95/46/EC. Eur-lex.europa.eu. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046.

[2] Google and Facebook tighten grip on us digital ad market. Emarketer.com, Sep 2017. https://www.emarketer.com/Article/Google-Facebook-Tighten-Grip-on-US-Digital-Ad-Market/1016494.

[3] ANDREOU, A., VENKATADRI, G., GOGA, O., GUMMADI, K. P., LOISEAU, P., AND MISLOVE, A. Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations. In *NDSS 2018, Network and Distributed Systems Security Symposium 2018, 18-21 February 2018, San Diego, CA, USA* (San Diego, ÉTATS-UNIS, 02 2018).

[4] CARRASCOSA, J. M., MIKIANS, J., CUEVAS, R., ERRAMILLI, V., AND LAOUTARIS, N. I always feel like somebody's watching me: Measuring online behavioural advertising. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies* (New York, NY, USA, 2015), CoNEXT '15, ACM, pp. 13:1–13:13.

[5] CASTELLUCCIA, C., KAAFAR, M.-A., AND TRAN, M.-D. Betrayed by your ads! In *International Symposium on Privacy Enhancing Technologies Symposium* (2012), Springer, pp. 1–17.

[6] DE PROTECCIÓN DE DATOS, A. E. The spanish dpa fines facebook for violating data protection regulations, 11 September 2017. http://www.agpd.es/portalwebAGPD/revista_prensa/revista_prensa/2017/notas_prensa/news/2017_09_11-iden-idphp.php.

[7] ERIKSON, E. H., AND ERIKSON, J. M. *The life cycle completed (extended version)*. WW Norton & Company, 1998.

[8] EU. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation), 27 April 2016. http://eur-lex.europa.eu/eli/reg/2016/679/oj.

[9] FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters 27*, 8 (2006), 861–874.

[10] FLEISS, J. L. Measuring nominal scale agreement among many raters. *Psychological bulletin 76*, 5 (1971), 378.

[11] FLEISS, J. L., LEVIN, B., AND PAIK, M. C. *Statistical methods for rates and proportions*. John Wiley & Sons, 2013.

[12] GONZÁLEZ CABAÑAS, J., CUEVAS, Á., AND CUEVAS, R. FDVT: Data Valuation Tool for Facebook users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, CO, USA, 2017), ACM, pp. 3799–3809.

[13] HAN, X., KHEIR, N., AND BALZAROTTI, D. Phisheye: Live monitoring of sandboxed phishing kits. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (New York, NY, USA, 2016), CCS '16, ACM, pp. 1402–1413.

[14] HONG, J. The state of phishing attacks. *Commun. ACM 55*, 1 (Jan. 2012), 74–81.

[15] KORPUSIK, M., COLLINS, Z., AND GLASS, J. Semantic mapping of natural language input to database entries via convolutional neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on* (2017), IEEE, pp. 5685–5689.

[16] LANDIS, J. R., AND KOCH, G. G. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[17] MIKOLOV, T., CHEN, K., CORRADO, G., AND DEAN, J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[18] MIKOLOV, T., SUTSKEVER, I., CHEN, K., CORRADO, G. S., AND DEAN, J. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (2013), pp. 3111–3119.

[19] MIKOLOV, T., YIH, W.-T., AND ZWEIG, G. Linguistic regularities in continuous space word representations. In *hlt-Naacl* (2013), vol. 13, pp. 746–751.

[20] NARASIMHAMURTHY, A. Theoretical bounds of majority voting performance for a binary classification problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*, 12 (2005), 1988–1995.

[21] OPINION, T., AND SOCIAL. Special eurobarometer 431 data protection, 2015. http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_431_en.pdf.

[22] PANCHENKO, A. Best of both worlds: Making word sense embeddings interpretable. In *LREC* (2016).

[23] PENNINGTON, J., SOCHER, R., AND MANNING, C. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (2014), pp. 1532–1543.

[24] RICCI, F., ROKACH, L., SHAPIRA, B., AND KANTOR, P. B. *Recommender Systems Handbook*, 1st ed. Springer-Verlag New York, Inc., New York, NY, USA, 2010.

[25] SPEICHER, T., ALI, M., VENKATADRI, G., RIBEIRO, F. N., ARVANITAKIS, G., BENEVENUTO, F., GUMMADI, K. P., LOISEAU, P., AND MISLOVE, A. Potential for discrimination in online targeted advertising.

[26] VENKATADRI, G., LIU, Y., ANDREOU, A., GOGA, O., LOISEAU, P., MISLOVE, A., AND GUMMADI, K. P. Privacy risks with Facebook's PII-based targeting: Auditing a data broker's advertising interface. In *S&P 2018, IEEE Symposium on Security and Privacy, 20-24 May 2018, San Francisco, CA, USA* (San Francisco, ÉTATS-UNIS, 05 2018).

[27] WEISCHEDEL, R., PALMER, M., MARCUS, M., HOVY, E., PRADHAN, S., RAMSHAW, L., XUE, N., TAYLOR, A., KAUFMAN, J., FRANCHINI, M., ET AL. Ontonotes release 5.0 ldc2013t19. *Linguistic Data Consortium, Philadelphia, PA* (2013).

[28] ZHU, W., ZENG, N., WANG, N., ET AL. Sensitivity, specificity, accuracy, associated confidence interval and roc analysis with practical sas implementations. *NESUG proceedings: health care and life sciences, Baltimore, Maryland 19* (2010).

Apéndice

A Excepciones del RGPD para el procesamiento de datos personales sensibles

A continuación, enumeramos las excepciones incluidas en el Artículo 9 del RGPD que permiten el procesamiento de información sensible. En las excepciones, el término interesado se refiere a los usuarios en el contexto de FB y el término responsable se refiere a FB en sí. Según nuestro conocimiento, ninguna de las excepciones se aplicaría a las preferencias publicitarias sensibles de FB.

(a) el interesado dio su consentimiento explícito para el tratamiento de dichos datos personales con uno o más de los fines especificados, excepto cuando el Derecho de la Unión o de los Estados miembros establezca que la prohibición mencionada en el apartado 1 no puede ser levantada por el interesado;

(b) el tratamiento es necesario para el cumplimiento de obligaciones y el ejercicio de derechos específicos del responsable del tratamiento o del interesado en el ámbito del Derecho laboral y de la seguridad y protección social, en la medida en que así lo autorice el Derecho de la Unión de los Estados miembros o un convenio colectivo con arreglo al Derecho de los Estados miembros que establezca garantías adecuadas del respeto de los derechos fundamentales y de los intereses del interesado;

(c) el tratamiento es necesario para proteger intereses vitales del interesado o de otra persona física, en el supuesto de que el interesado no esté capacitado, física o jurídicamente, para dar su consentimiento;

(d) el tratamiento es efectuado, en el ámbito de sus actividades legítimas y con las debidas garantías, por una fundación, una asociación o cualquier otro organismo sin ánimo de lucro, cuya finalidad sea política, filosófica, religiosa o sindical, siempre que el tratamiento se refiera exclusivamente a los miembros actuales o antiguos de tales organismos o a personas que mantengan contactos regulares con ellos en relación con sus fines y siempre que los datos personales no se comuniquen fuera de ellos sin el consentimiento de los interesados;

(e) el tratamiento se refiere a datos personales que el interesado ha hecho manifiestamente públicos;

(f) el tratamiento es necesario para la formulación, el ejercicio o la defensa de reclamaciones o cuando los tribunales actúen en ejercicio de su función judicial;

(g) el tratamiento es necesario por razones de un interés público esencial, sobre la base del Derecho de la Unión o de los Estados miembros, que debe ser proporcional al objetivo perseguido, respetar en lo esencial el derecho a la protección de datos y establecer medidas adecuadas y específicas para proteger los intereses y derechos fundamentales del interesado;

(h) el tratamiento es necesario para fines de medicina preventiva o laboral, evaluación de la capacidad laboral del trabajador, diagnóstico médico, prestación de asistencia o tratamiento de tipo sanitario o social, o gestión de los sistemas y servicios de asistencia sanitaria y social, sobre la base del Derecho de la Unión o de los Estados miembros o en virtud de un contrato con un profesional sanitario y sin perjuicio de las condiciones y garantías contempladas en el apartado 3;

(i) el tratamiento es necesario por razones de interés público en el ámbito de la salud pública, como la protección frente a amenazas transfronterizas graves para la salud, o para garantizar elevados niveles de calidad y de seguridad de la asistencia sanitaria y de los medicamentos o productos sanitarios, sobre la base del Derecho de la Unión o de los Estados miembros que establezca medidas adecuadas y específicas para proteger los derechos y libertades del interesado, en particular el secreto profesional,

(j) el tratamiento es necesario con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos, de conformidad con el artículo 89, apartado 1, sobre la base del Derecho de la Unión o de los Estados miembros, que debe ser proporcional al objetivo perseguido, respetar en lo esencial el derecho a la protección de datos y establecer medidas adecuadas y específicas para proteger los intereses y derechos fundamentales del interesado.

B Resolución de la AEPD asociada a la sanción a FB

En este apéndice, enumeramos los principales elementos incluidos en la resolución de la AEPD asociada con la multa de 1.2M€impuesta a FB por violar la regulación española de protección de datos.

- *La Agencia señala que la red social recopila, almacena y utiliza datos, incluidos datos especialmente sensibles, con fines publicitarios sin obtener consentimiento.*
- *Los datos sobre ideología, sexo, creencias religiosas, preferencias personales o actividad de navegación se recopilan directamente, a través de la interacción con sus servicios o desde páginas de terceros, sin informar claramente al usuario sobre cómo y para qué fines se utilizarán esos datos.*
- *Facebook no obtiene un consentimiento inequívoco, específico e informado de los usuarios para procesar sus datos, ya que la información que ofrece no es clara.*
- *Los datos personales de los usuarios no se eliminan totalmente cuando ya no son útiles para el propósito*

para el cual fueron recopilados, ni cuando el usuario solicita explícitamente su eliminación.

- *La AEPD declara la existencia de dos infracciones graves y una muy grave de la Ley de Protección de Datos e impone a Facebook una sanción total de 1.200.000 euros.*

- *La AEPD es parte de un grupo que junto con las Autoridades de Bélgica, Francia, Hamburgo (Alemania) y los Países Bajos, también iniciaron sus respectivos procedimientos de investigación a la compañía.*

C Términos de uso y política de anuncios en FB

Los usuarios de FB aceptan las Condiciones de Uso de Facebook[22] al abrir una cuenta de FB. Este es el documento de entrada donde los usuarios están informados de lo que FB está haciendo con sus datos personales. Sin embargo, para comprender mejor los detalles relacionados con la administración de datos de FB, los usuarios pueden dirigirse a otro documento denominado Política de Datos.[23] Encontramos tres secciones muy relevantes para nuestra investigación en el documento de Términos de Uso:

Sección 16. Disposiciones especiales aplicables a usuarios fuera de los Estados Unidos. Esta sección incluye la siguiente cláusula *"Usted acepta que sus datos personales sean transferidos y procesados en los Estados Unidos."* Mientras esto otorga a FB suficiente permiso para procesar y almacenar datos personales, el RGPD y las regulaciones de protección de datos anteriores en algunos países de la UE establecen una clara diferencia entre los datos personales y *"especialmente protegidos"* o *"sensibles"* . Según nuestro conocimiento, FB no obtiene un permiso explícito específicamente para procesar y almacenar datos personales sensibles.

Sección 9. Acerca de los anuncios y otro contenido comercial servido por Facebook. En esta sección, se informa a los usuarios que FB puede usar su información, nombre, imagen, etc. con fines publicitarios y comerciales.

Sección 10. Disposiciones especiales aplicables a los anunciantes. Los anunciantes son dirigidos a dos documentos más: Términos de Anuncios[24] (no muy relevante para nuestra investigación) y Políticas Publicitarias.[25] El último documento incluye 13 secciones de las cuales la Sección 4.12 [26] (4-Contenido prohibido; 12- Atributos personales) es muy relevante para nuestro artículo. La Sección 4.12 dice: *"Los anuncios no deben contener contenido que afirme o implique atributos personales. Esto incluye afirmaciones o implicaciones directas o indirectas sobre la raza, el origen étnico, la religión, las creencias, la edad, la orientación sexual o las prácticas de una persona, la identidad de género, la discapacidad, la afección médica (incluida la salud física o mental), el estado financiero, la afiliación a un sindicato, antecedentes penales o nombre."*. En las Políticas de Publicidad se proporcionan ejemplos de qué contenido está permitido y qué contenido está prohibido.

D Spacy

Spacy es un paquete gratuito de código abierto para operaciones avanzadas de PLN. Spacy ofrece múltiples funciones de PLN como extracción de información, comprensión del lenguaje natural, aprendizaje profundo de texto, análisis de similitud semántica, etc., que se llevan a cabo a través de diferentes modelos predefinidos. Para llevar a cabo nuestro análisis, aprovechamos la característica de "similitud" de Spacy que permite comparar dos palabras o texto corto que proporciona un valor de similitud semántica que oscila entre 0 (el más bajo) y 1 (el más alto). Esta característica calcula la similitud utilizando el método denominado GloVe (vectores globales para la representación de palabras) [23]. Los GloVe son representaciones de significados multidimensionales de palabras calculadas usando word2vec [17][18][19].

Los vectores de palabras espaciales se entrenan utilizando un gran corpus de texto que incorpora un rico vocabulario. Además, Spacy también tiene en cuenta el contexto para definir la representación de una palabra, lo que le permite a Spacy identificar mejor su significado considerando las palabras circundantes. Spacy ofrece diferentes modelos para optimizar el cálculo de similitud semántica. Hemos elegido el modelo *en_core_web_md*[27] porque optimiza el análisis de similitud entre palabras y oraciones cortas, que coincide con la naturaleza de los nombres de preferencias publicitarias. El modelo elegido es una red neuronal convolucional (CNN, por sus siglas en inglés) multitarea entrenada en OntoNotes [27] con vectores GloVe que a su vez están

[22] `https://www.facebook.com/terms.php` (consultada el 19 de diciembre de 2017)

[23] `https://www.facebook.com/about/privacy/` (consultado el 19 de diciembre , 2017)

[24] `https://www.facebook.com/legal/self_service_ads_terms` (consultado el 19 de diciembre de 2017)

[25] `https://www.facebook.com/policies/ads/` (consultado el 19 de diciembre de 2017)

[26] `https://www.facebook.com/policies/ads/prohibited_content/personal_attributes` (consultado el 19 de diciembre de 2017)

[27] `https://spacy.io/models/en#en_core_web_md`

entrenados en Common Crawl.[28] Common Crawl es un repositorio de código abierto para crawlear datos. El modelo utiliza vectores de palabras, vectores de token específicos del contexto, etiquetas POS (Part-Of-Speech), análisis de dependencia y reconocimiento de palabras.

E Campañas llevadas a cabo respetando los términos de anuncios de FB

Figure 11: FDVT anuncio

Figure 12: FDVT anuncio

Las Figuras 11 y 12 muestran los dos anuncios que hemos utilizado en nuestras campañas. Estos anuncios hacen referncia a nuestra extensión de navegador FDVT y, por lo tanto, no incluyen contenido que afirme o implique atributos personales. De hecho, la página de destino donde se redirigió a los usuarios en caso de que hicieran clic en alguno de estos anuncios es la página web del proyecto FDVT.[29]

En los experimentos, no registramos ninguna información de los usuarios que hicieron clic en los anuncios y visitaron nuestra página de destino. La única información que utilizamos en este documento es la que proporciona FB a través de los informes que ofrece a los anunciantes relacionados con sus campañas publicitarias.

[28] http://commoncrawl.org/
[29] https://www.fdvt.org/