



Facultad de Ciencias Económicas y Empresariales

EL BIG DATA Y LA HUELLA DIGITAL: LA IMPORTANCIA DE LOS DATOS Y CÓMO SON UTILIZADOS POR LAS EMPRESAS

Autor: Gerardo Preciado Vidal-Aragón
Director: Lucía Barcos Redín

MADRID | Junio 2019

Resumen

En el presente trabajo se pretende explicar qué es el Big Data y en qué se diferencia del Business Intelligence, así como el concepto de huella digital y porqué es importante en la actualidad. Nos encontramos en un mundo en el que generamos datos constantemente. Veremos cómo el Big Data y la huella digital son elementos fundamentales en la toma de decisiones de determinadas empresas con sus clientes y cómo, indirectamente y sin que nos demos cuenta, pueden afectar los datos a las personas.

- Palabras clave: Big Data, Business Intelligence, datos, huella digital, análisis, decisiones, privacidad.
-

Abstract

The present project's aim is to explain what Big Data is and how it differs from Business Intelligence, as well as the concept of fingerprint and why it is important today. We are in a world where we generate data constantly. We will see how Big Data and the digital fingerprint are fundamental elements in the making decision of certain companies with their clients and how, indirectly and without us realizing it, people can be affected by data.

- Key words: Big Data, Business Intelligence, data, digital footprint, analysis, decisions, privacy.

Índice

1. Introducción.....	5
2. Que es el Big Data.....	7
2.1. Concepto y características.....	7
2.2. Historia y evolución.....	10
2.3. Business Intelligence y Big Data.....	13
3. Diversas aplicaciones del Big Data	16
4. Huella digital.....	18
4.1. Qué es la huella digital.....	18
4.2. Nuestros datos en las empresas.....	23
5. Cómo afectan a las empresas las leyes de protección de datos.....	39
6. Resumen y conclusiones.....	41
7. Bibliografía.....	43

Abreviaturas

BD	Big Data
BI	Business Intelligence
PDF	Portable Document Format
HTML	HyperText Markup Languaje
WWW	World Wide Web
HD	Huella Digital
IP	Internet Protocol
FB	Facebook
RGPD	Reglamento General de Protección de Datos
LO	Ley Orgánica
LOPDGDD	Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales

1. Introducción

El mundo está en constante cambio y nosotros debemos adaptarnos a él. La gran revolución digital ha producido que estemos generando datos en todo momento y estos son almacenados y analizados por terceros. Los datos se han convertido en un activo de un valor incalculable para entidades del gobierno y empresariales.

En este trabajo definiremos qué es el Big Data y la huella digital que dejamos en la red. La aplicación del tratamiento de datos masivos es cada vez de mayor relevancia tanto para las empresas como para las propias personas. El objetivo final de una empresa es, generalmente, crear valor, y para ello es importante definir un plan estratégico que analice los procesos de creación y transformación de valor en todas sus formas (para los accionistas, para la empresa, para los clientes, para los trabajadores, para el sector, etc.).

Las empresas buscan maximizar y profundizar en el análisis de estos datos para mejorar la actividad de la empresa, optimizar recursos y elaborar planes estratégicos que creen valor para la misma. Veremos qué es Business Intelligence y sus herramientas, pues estas permiten, basándose en la predicción de comportamientos de hechos reales, realizar análisis predictivos y avanzados que ayudan a las empresas en la toma de decisiones. Actualmente, es imprescindible para las empresas tener la gran habilidad de poder transformar datos en información útil para estas.

El Big Data ha experimentado una gran evolución desde sus inicios. Hoy día generamos datos desde el momento en el que cogemos el móvil al levantarnos. Los móviles dejan un rastro de nuestra localización en cada momento, los sitios donde compramos, lo que buscamos por Internet o las páginas a las que accedemos, datos sanitarios personales y comentarios en redes sociales. En resumen, todo aquello que hacemos genera un registro, una huella, que queda almacenada en una base de datos. Este rastro puede ser utilizado por determinadas organizaciones en su beneficio según sean sus intereses. Un ejemplo de ello es Google que, como veremos más adelante, utiliza nuestros datos para la elaboración de perfiles individuales en función de las búsquedas en la web y los gustos de cada uno. A lo largo de este trabajo veremos algunos ejemplos en los que los datos han servido para facilitar a las empresas la toma de decisiones y algunos han perjudicado a los usuarios.

Estructura del trabajo.

En cuanto a la estructura del trabajo, se ha comenzado introduciendo qué es el Big Data y su historia y evolución. Veremos cómo esta realidad que aún está por explorar es de gran utilidad en la aplicación práctica del mundo de la empresa. También trataremos qué es Business Intelligence y la utilidad que tiene, así como su relación con el Big Data. Posteriormente hemos desarrollado algunas de las diversas aplicaciones del Big Data en el mundo en el que vivimos. El siguiente punto para tratar es el concepto de huella digital, dentro de este marco veremos en qué consiste la huella digital y sus principales características, así como la relevancia de la información personal para las empresas. También veremos cómo utilizan nuestros datos empresas como Google, Facebook y Amazon y cuánta información tienen sobre nosotros. Por último, y no por ello menos importante, veremos cómo las leyes vigentes protegen la privacidad de las personas en cuanto a los datos que generan.

2. Qué es el Big Data

2.1. Concepto y características

Big Data es un término que se ha popularizado en los últimos años. Nació con el fin de cubrir necesidades como el almacenamiento y tratamiento de grandes volúmenes de datos. A diario se producen cantidades ingentes de datos por medio de una inmensa cantidad de dispositivos y, lo que para algunos son simples clics en páginas o comentarios en redes, para otros es información que puede ser de gran utilidad. Cuando hablamos de BD no solo hablamos de gran volumen de datos, también nos referimos al proceso de recogida de estos, el almacenamiento y el análisis, todo aquello por lo que se obtiene valor. La sociedad de la información se ve a simple vista, hoy día quién no tiene un teléfono móvil o un ordenador, se conecta a la red para buscar información de interés o realiza compras por internet. La aceleración de la digitalización, impulsada por el aumento de las nuevas tecnologías, ha dado lugar a un mundo de interrupciones sin precedentes y una creciente ambigüedad, con la que viene la incertidumbre y el riesgo, pero también la oportunidad. En este nuevo mundo de datos, las empresas necesitan encontrar cuál es su propósito.

La primera definición que daremos es que el Big Data es un avance tecnológico que da un nuevo enfoque a la toma de decisiones y al entendimiento, y que es utilizado para describir las enormes cantidades de datos. (Ricardo Barranco, 2012).

Otra definición muy significativa es la de la consultora Gartner¹ sobre las características que marcan la diferencia “Conjunto de datos de gran volumen, alta velocidad y/o gran variedad de activos de información que exigen formulas innovadoras y rentables de procesamiento de la información y que permiten una visión mejorada, la toma de decisiones y la automatización de procesos”. (Gartner, 2015). Gartner considera que la esencia del Big Data es todo aquello que se puede conseguir aprovechando el potencial de los datos.

¹ Gartner: empresa consultora y de investigación de las tecnologías de la información con sede en Stamford, Connecticut, Estados Unidos.

De todas las definiciones cabe destacar la de Gartner sobre las tres características más importantes (las tres V's): volumen, velocidad y variedad. No obstante, la empresa IBM incluye una cuarta (veracidad). A continuación, definiremos dichas características:

- Volumen: hace referencia a la cantidad de información que se genera constantemente en nuestro entorno. Las empresas y las personas generan datos en todo momento: foros, aplicaciones, redes sociales. Suelen ser datos desestructurados pero que con un análisis de estos puedes ser potencialmente valiosos.
- Velocidad: es la frecuencia con la que los datos son obtenidos, almacenados y analizados. La velocidad de creación de datos ha aumentado exponencialmente debido a las continuas interconexiones entre dispositivos. La rapidez de transformación de datos puede ser un factor clave para el posicionamiento de determinadas empresas en el mercado.
- Variedad: es referida a las diversas fuentes y formas donde se registran los datos. Podemos encontrarnos con datos estructurados, semiestructurados y no estructurados, con los que es más complicado trabajar, si bien, es por eso por lo que tienen más riqueza y valor
- Veracidad: se entiende como el sesgo y la alteración de los datos. Hay que estar seguros de que la información es válida y que los datos obtenidos son significativos y están relacionados con el objetivo por el que se quieren analizar.

Los datos que se generan hoy en día son una fuente enorme de información que antes no estaba a nuestro alcance. La mayoría son muy apreciados por las empresas, ya que dan una información de los clientes muy valiosa para la mejora y el desarrollo de la compañía. A través de análisis automáticos se transforman en información completamente nueva que ayuda a las empresas a tomar medidas inteligentes. Éstas aprovechan los datos y los análisis para las decisiones y procesos que ayudará a incrementar el valor empresarial.

De las características mencionadas anteriormente, para algunas empresas será más importante el volumen de datos, para otras será la velocidad y, para otras veracidad o variedad. Todo depende de la finalidad y el objetivo de la empresa, por ejemplo: las empresas que tengan un gran volumen se interesarán más en capturar información, almacenarla y utilizarla para sus procesos de negocio, si bien, hay algunas que no necesitan tanto el volumen sino trabajar en tiempo real y a gran velocidad.

Existen tres tipos de datos que mostraremos a continuación:

- Estructurados: archivos de tipo texto que se encuentran en la mayoría de las bases de datos y se encuentran en filas y columnas con título. Se almacenan en bases de datos relacionales que cumplen con el modelo relacional² (más utilizado para implementar las bases de datos ya planificadas). Se pueden almacenar en tablas y tienen una clara definición de longitud y formato.
- No estructurados: datos internos que no tienen estructura identificable. Conglomerado de objetos que carecen de valor hasta ser organizados y analizados. No poseen un formato específico que permita almacenarlos de forma tradicional. Entre ellos son comunes documentos Microsoft Office Word, archivos en formato PDF, correo electrónico, audio, etc.
- Semiestructurados: son una mezcla de los dos anteriormente mencionados, no presentan una estructura definida, pero sí una organización de los metadatos donde se describen los objetos. Poseen ciertos patrones que los describen y dan información sobre las relaciones entre los mismos. Como ejemplo el HTML³.

² Modelo relacional: es un modelo de datos basado en la lógica de predicados (diseñado para estudiar la inferencia en lenguaje de primer orden) y en la teoría de conjuntos (rama de la lógica matemática que estudia relaciones de los conjuntos).

³ HTML: lenguaje de programación que se utiliza para el desarrollo de páginas de Internet.

2.1. Historia y evolución

Podríamos poner los inicios en Alan Turing⁴ que es considerado el “padre” de la computación. Hacia los años 30 no existía ninguna máquina ni aparato capaz de ejecutar una tarea y cambiar a otra de forma autónoma, debía hacerse manualmente. Sin embargo, fue durante la Segunda Guerra Mundial cuando Turing lideró un equipo para diseñar una máquina que serviría para descifrar Enigma (máquina de cifrado del ejército nazi). Por ello se considera a Turing como el que puso en marcha el Big Data. A continuación, veremos cuáles son los hitos más importantes:

Big Data desde los 50 a los 80:

- 1940: en esta época los únicos lugares de almacenamiento de datos eran las bibliotecas.
- 1956: Fritz Rudolf Güntsch desarrolla la memoria virtual que trataba el almacenamiento de datos, finito o infinito.
- 1962: William C. Dersch presenta un Sistema capaz de comprender la voz, Shoebox⁴.
- 1966: a finales de los 60 las empresas empiezan a desarrollar mejores sistemas de computación. Esto permite a las empresas automatizar sistemas internos.
- 1980: es en esta fecha cuando I.A. Tjomsland comentó la Ley de Parkinson, esta dice: “el trabajo se expande hasta llenar el tiempo disponible para que se termine”. Parafraseando se podría decir: “los datos se expanden hasta llenar el espacio disponible para el almacenamiento”

⁴ Alan Turing: matemático, lógico y científico de la computación. Proporcionó una influyente formalización de los conceptos de algoritmo y computación; siglo XX.

Desde el nacimiento de la WWW e Internet hasta hoy:

- 1989: nace la World Wide Web por Tim Berners-Lee, Sistema que distribuye documentos a Internet.
- 1992: Crystal Reports⁵ desarrolla un avance, la primera base de datos, lo cual facilita el trabajo a las empresas gracias a un trabajo más informatizado y las impulsa hacia el crecimiento en el ámbito de la programación.
- 1997: fue la primera vez que se utilizó el término Big Data. Investigadores de la NASA como Michael Cox y David Ellsworth sostenían que el enorme crecimiento que habían experimentado los datos iba a suponer un problema para los sistemas informáticos de esa época.
- 2009-2011: nacen empresas dedicadas a gestionar servicios como el almacenamiento y centralización de datos.
- 2012: en esta fecha el Big Data fue utilizado por Barack Obama con fines políticos para conocer opiniones de ciudadanos. Fue la primera vez que se utilizaba en el ámbito político.
- 2014: “Location Intelligence”: consistía en la obtención de datos de geolocalización para obtener valor.
- 2015: Smart Cities: surge por la evolución de las Ciudades Digitales nacidas en 2004 en España. Comienzan a aparecer por la evolución y tratamiento de los datos.
- 2020: futuro del Big Data: los expertos apuntan un aumento de la generación de datos del 4300% al año.

⁵ Crystal Reports: es una aplicación empleada para el diseño de informes desde una amplia base de datos.

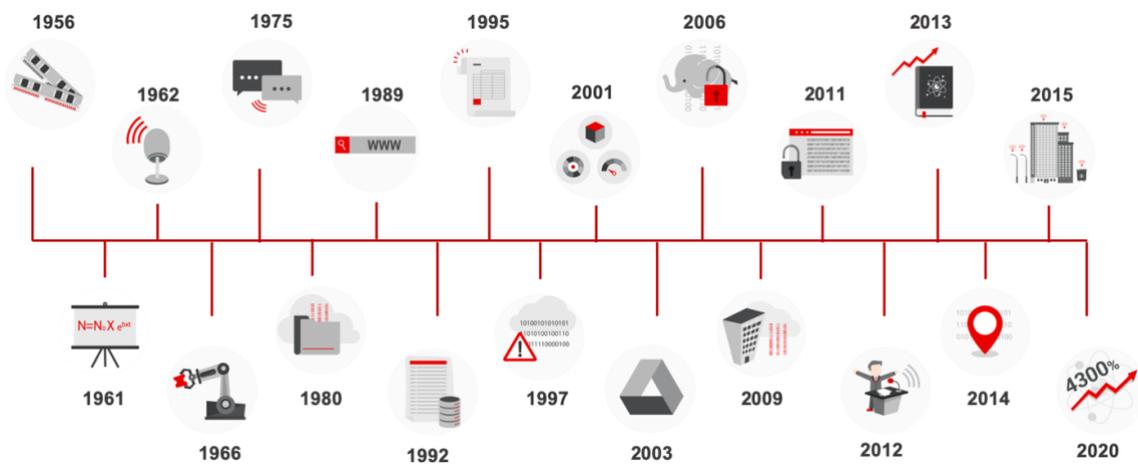


Ilustración 1 Cronología de la evolución del Big Data (Forbes, Winshuttle, Wikipedia, Gartner, IBM)

2.3. Business Intelligence y Big Data

Con el rápido desarrollo de nuevas tecnologías el entorno empresarial y diversos sectores de la sociedad están experimentando profundos cambios. Dentro del mundo de los negocios se han podido crear infraestructuras más complejas, aumentar el nivel de comunicación mediante la creación de redes de empresas, alianzas y nuevos negocios. Lo que esto nos permite es acumular los datos de la empresa y utilizarlos para mejorar la productividad y desarrollar nuevos modelos de negocio o mejorar los existentes. Bajo este nuevo paradigma la toma de decisiones debe realizarse con mayor rapidez y precisión, y para conseguirlo se han de utilizar herramientas diseñadas para este objetivo. Business Intelligence trata de conocerse mejor a sí mismo, es decir, a la empresa que lo implementa a través de la creación de modelos de datos que tengan en cuenta métricas y objetivos cuya finalidad sea la mejora de toma de decisiones. Permite mejorar la competitividad de la empresa mediante las siguientes acciones: recopilación, análisis y comprensión de los datos; la evolución de los datos permite realizar predicciones; colaboración entre departamentos; y como última acción, decisión, por la cual se lleva a cabo el plan establecido.

“La Inteligencia de Negocio (BI) es un término genérico que incluye las aplicaciones, la infraestructura y las herramientas, y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y rendimiento”. (Gartner, 2015).

Este concepto permite convertir de manera inteligente los datos almacenados por la empresa en información útil y valiosa utilizando técnicas, herramientas y tecnología. El objetivo es generar conocimientos para mejorar la toma de decisiones. Este tipo de herramientas permite obtener un mayor conocimiento de los factores que influyen en el modelo de negocio de la organización y según su complejidad podemos clasificarlas en:

- Consultas e informes: presentan la información de una manera predeterminada. Los usuarios pueden filtrar la información a través de parámetros establecidos y pueden decidir criterios de ordenamiento.

- Cubos OLAP: permiten realizar consultas online mediante un modelo parecido a las tablas dinámicas. Un cubo es una estructura de datos multidimensional, su facilidad de uso y su rapidez pueden ser unas ventajas de la utilización de esta herramienta, sin embargo, el modelo de negocio no siempre se adapta a un modelo jerárquico.
- Data mining: minería de datos en español, su objetivo es identificar patrones y relaciones que no son evidentes a través de los métodos tradicionales de análisis de información.
- Tableros de control: se suelen representar de forma visual mediante la combinación de gráficos interactivos y tablas de datos. Es un cuadro de mando para realizar un seguimiento de los procesos operativos para detectar situaciones críticas.
- Notificaciones: consiste en la ejecución de alertas y alarmas a través de software que cuando determinados KPI⁶ muestran valores que difieren de los objetivos perseguidos se activan para avisar al responsable.
- Predicción: este tipo de herramientas permite extraer conclusiones sobre eventos futuros por medio de la aplicación de métodos estadísticos y matemáticos.

⁶ KPI: “Key Performance Indicator”, los indicadores clave de rendimiento dan una referencia del rango en el que debería situarse una empresa cuando alcanza sus objetivos. También existen los KGI (Key Goal Indicator) los cuales establecen una serie de medidas para controlar si un proceso TIC ha alcanzado sus objetivos.

Podríamos decir que Big Data es un mundo sobre el que todavía se tiene que investigar y Business Intelligence es la realidad más cercana a las empresas, si bien, ambos se utilizan para conseguir el mismo objetivo: obtener valor para el negocio. Hay que decir que ambos utilizan distinto modo de ejecución, pero se complementan. En cuanto a Big Data, los datos aparecen en cualquier entorno y en formatos diversos, por otro lado, BI profundiza en los conjuntos de datos de los que dispone la compañía y sus herramientas permiten acceder a conjuntos ya clasificados de estos de manera que se puedan analizar, sus conclusiones proporcionarán una información detallada a los usuarios.

3. Diversas aplicaciones del Big Data

Conocimiento de los clientes

Actualmente los departamentos de marketing y ventas son las áreas donde se utiliza más el Big Data ya que los datos son utilizados para tener una visión más profunda de los clientes y sus comportamientos, es decir, un conocimiento total de estos. El objetivo que persiguen es el entender el comportamiento del cliente a través de la creación de modelos predictivos analizando textos y datos.

Optimizando los procesos de negocio

Cada vez se está aplicando más el Big Data en las empresas para mejorar la optimización de procesos. Gracias a los datos recogidos de redes sociales y búsquedas en Internet las empresas están optimizando su stock por la mejora de las predicciones realizadas sobre estos datos. Debido a la aparición de sensores de identificación y un mejor posicionamiento geográfico, los cuales permiten una visualización del tráfico en tiempo real (entre otros), la cadena de suministro y las rutas de reparto han experimentado grandes avances.

Mejora del rendimiento personal y deportivo.

No solo se pueden beneficiar las empresas y los gobiernos de los avances en el análisis de datos, nosotros también. Existen dispositivos que registran el consumo de calorías de una persona, nivel de actividad física, datos médicos y otros muchos datos que complementan la vida de las personas, como por ejemplo los smart watches. Además, esto puede ser un punto de mejora del rendimiento deportivo. Actualmente, la mayoría de los deportistas utilizan procedimientos de análisis de datos para ver los puntos donde flaquean y así poder incrementar el rendimiento. Un ejemplo es el empleo de sensores en balones de baloncesto y palos de golf; de esta manera se analiza el recorrido y se obtiene una retroalimentación que será utilizada para una posterior mejora del jugador. Cabe destacar la utilización de Slam Tracker⁷ en tenis.

⁷ Slam Tracker: herramienta utilizada en tenis a través de la cual, mediante una serie de análisis predictivos, compara información del tenista, el momento del partido y el sentimiento social.

Mejora de la Salud Pública

Otra de las aplicaciones se centra en la salud. En algunos hospitales el tratamiento de datos masivos se emplea para la monitorización de bebés, gracias a la grabación de los latidos del corazón y la respiración se crean algoritmos que pueden prever la aparición de alguna enfermedad infecciosa hasta 24 horas antes. Por otro lado, integrando los historiales médicos de los pacientes y según lo que publican las personas en sus perfiles también es utilizado para el control de epidemias o brotes de gripe.

Optimizando el rendimiento de máquinas y dispositivos

El análisis de los datos ha contribuido a que las máquinas y los dispositivos comiencen a ser cada vez más autónomos e inteligentes. Un ejemplo de ello es el coche autopilotado de Google; coche equipado con todo tipo de necesidades como GPS, conexión a la red y un gran número de sensores, que permiten que el coche circule por la vía pública sin un conductor humano. También destaca la empresa israelí Otonomo, encaminada a liderar el campo de la recopilación y procesamiento de los datos que producen los coches.

Recursos Humanos

Cada vez son más las compañías que utilizan una estrategia basada en métricas para captar a los nuevos empleados. A través de un análisis de los diversos candidatos se puede mejorar la gestión de recursos humanos y atraer a nuevos talentos a la empresa, además de conseguir una mayor alienación con los objetivos del negocio. Permite saber qué perfiles son necesarios para cada una de las áreas a las que se dedica el negocio y realizar una mejora de la segmentación de este.

Ámbito económico

En el ámbito financiero el Big Data es principalmente utilizado en actividades que tienen relación con High-Frequency Trading. Se pueden realizar análisis en un tiempo muy reducido sobre las decisiones de compra venta de valores en el mercado. En este sector, destaca Social Media Analytics que mide el impacto económico de una empresa y el análisis de la imagen de la marca en el mundo digital. Además, esta herramienta permite monitorizar la evolución de las acciones.

4. La Huella Digital

4.1. Definición y concepto

El rastro que dejamos al utilizar un dispositivo tecnológico es lo que conocemos como Huella Digital. Cada uno de nosotros está constantemente utilizando el teléfono móvil, el ordenador o una “Tablet”. Podríamos decir que es la suma de lo que publicamos, compartimos y lo que publican sobre nosotros. A veces es muy sencillo, basta con poner nuestro nombre en un buscador de internet y veremos qué dice sobre nosotros. En el mundo online, hay que preocuparse tanto de lo que uno publica voluntariamente como de lo que los demás publican de nosotros. Podemos distinguir dos tipos:

- Huella digital activa: cada vez que ejecutamos una acción o compartimos algo online, es decir, si somos nosotros los que ejecutamos la acción.
- Huella digital pasiva: en el caso en que sean otros los que publican sobre nosotros.

Es de vital importancia cuidarla, pues es una de las fuentes de información más usada por las empresas, ya sea para conocer mejor a algún candidato a un puesto en la empresa o para conocer mejor a sus clientes y establecer una estrategia orientada a ellos. En las empresas afecta en gran medida a la reputación, factor fundamental para obtener y retener clientes.

Estamos acostumbrados a mirar todo tipo de información en internet, sin embargo, también nos están mirando a nosotros. Cada vez que buscamos información en la web, cuando compartimos algo en las redes sociales, enviamos un mensaje instantáneo o enviamos un correo electrónico, vamos dejando un rastro de información. Las huellas digitales pueden ofrecer comodidad como por ejemplo a la hora de iniciar una sesión en una aplicación, podemos tener guardado nuestro usuario y contraseña. La elaboración de un perfil se basa en recopilar información y búsquedas de los usuarios y extrae datos de configuración de las aplicaciones, geolocalización, programas instalados, etc. Si combinamos toda esa información y los datos extraídos se confeccionará lo que llamamos Huella Digital única de cada dispositivo y, por lo tanto, es diferente para cada usuario.

La mayoría de nosotros no nos damos cuenta de que al subir un comentario o una foto a las redes sociales estamos perdiendo parte de nuestra privacidad, pensamos que es una simple imagen que van a ver los usuarios que tenemos confirmados como amigos y no la puede ver nadie más. Si bien, es mas probable que no nos demos cuenta en las huellas que creamos por defecto: utilizar un buscador, realizar compras en internet o habilitar los servicios de localización. Las huellas digitales revelan una gran cantidad de información sobre nosotros, nuestros gustos y hábitos. Todos esos datos tienen un gran valor comercial, son monetizados por organizaciones con las que no tenemos ninguna relación y sobre las que no tenemos ningún control. Pocas personas son conscientes de lo íntimas que son nuestras huellas digitales y de la frecuencia con que los datos son compartidos con terceros. Incluso si pensamos que no tenemos nada que ocultar, no toda nuestra información es apropiada para todos los públicos. Nuestras huellas digitales pueden ser sacadas de contexto y malinterpretadas, por ejemplo: buscamos en internet acerca del terrorismo y nos documentamos sobre los grupos terroristas más expandidos por el mundo, esto puede dar a entender e interpretar aspectos y disciplinas de esa persona equivocados ya que puede que únicamente esté realizando un trabajo tratando ese tema; otro ejemplo es si estamos por la noche con los amigos y una sola persona paga la cuenta y mas adelante paga con la misma tarjeta en una gasolinera, se podría entender por un tercero que ha cogido el coche bajo efectos del alcohol, aunque no haya consumido absolutamente nada.

Un ejemplo es el transporte público (las bicicletas, el metro, el autobús), en general, puede dar una idea de nuestro recorrido en todo momento ya que, actualmente, cuando queremos movernos de un lado a otro es necesario utilizar una tarjeta de transporte a través de la cual quedan registrados nuestros movimientos. Por otro lado, con las nuevas empresas de alquiler de motos y coches eléctricos como medio de transporte, también estamos generando una enorme fuente de datos. Cada vez que un usuario alquila una moto para trasladarse, desde el momento en que la reserva hasta que la aparca se sabe donde ha estado, el recorrido que ha realizado y el lugar donde se dirigía. Esto puede dar una idea de dónde vives, donde trabajas, por dónde te sueles mover y, realmente, no somos conscientes de la información que ofrecemos, solo vemos que podemos trasladarnos con mayor rapidez y comodidad.

Si bien, aquello con lo que generamos más información de geolocalización es el teléfono. Con el móvil se pueden obtener registros diarios de los puntos en los que ha estado una persona en todo momento y cuánto tiempo ha estado en cada uno. Conociendo la localización de las personas se puede analizar su movilidad. Una vez que esto se conoce se pueden elaborar modelos predictivos para cuantificar el volumen de movimientos.

Todos tenemos información que solo queremos compartir en determinados contextos, puede que tengamos la confianza suficiente para compartir determinada información con una empresa, pero no que esta empresa la divulgue a otras. Con los mecanismos de seguimiento que existen en la actualidad la mayoría de los usuarios de internet no tiene el control necesario para evitar este tipo de difusión. La reputación es un juicio social y personal y nadie tiene derecho a compartirlo.

Una de las obtenciones de datos más frecuentes es a través de las aplicaciones. A diario se descargan numerosas aplicaciones que son gratuitas en los dispositivos tecnológicos, pero, realmente estamos pagando con nuestros datos. Lo que quiero decir es que al descargarnos una aplicación permitimos el acceso a más como pueden ser contactos para poder compartir contenido con nuestros amigos, almacenamiento para poder guardar elementos o incluso al teléfono si se trata de un calendario o una agenda.

Aplicación de ocio Clash Royale

Un ejemplo de esto es el juego Clash Royale, tiene 354 millones de descargas y 27.4 millones de jugadores de 187 países han participado en el torneo mundial Clash Royale Crash Championship. ¿Por qué este ejemplo? Como bien sabemos una gran cantidad de aplicaciones se descargan a diario en los dispositivos, entre ellas, aplicaciones de ocio. Este es uno de los juegos más descargados en el mundo y apenas nos fijamos en la política de privacidad, de la cual hablaremos de algunas características a continuación:

Información que se proporciona directamente:

- Nombre y correo electrónico.
- Nombre de usuario y contraseña.

- Información de perfil (fotografía).
- Registros de los chats del juego.
- Información en caso de que se pierda la cuenta.

Si nos damos cuenta nos dice: información que se proporciona **directamente**. Con esto somos nosotros los que proporcionamos la información conscientemente y estamos de acuerdo en que la tengan. Es requerida para poder empezar a jugar.

Información que recopilan de forma automática:

- Cuenta y progreso del juego.
- Dirección IP y la ID del dispositivo donde se juega.
- Información del dispositivo.
- Cookies⁸.
- Ubicación general.
- Geoubicación (con su consentimiento).
- Información sobre el uso del Servicio.

Toda esta información se recopila de forma **automática**, es decir, una vez que aceptamos la política de privacidad la aplicación tiene permiso para obtener el nombre del dispositivo, sistema operativo, idioma del explorador, ubicación donde se ha descargado o en todo momento cada vez que se accede al juego.

Podríamos pensar que toda esta cantidad de información que están obteniendo de nosotros es solamente para esa empresa, en este caso Supercell, empresa finlandesa que ha obtenido los últimos años unos ingresos de ocho mil millones de dólares con los juegos podríamos pensar que toda esta cantidad de información que están obteniendo de nosotros

⁸ Cookies: archivos creados por sitios web que contienen pequeñas cantidades de datos y se envían entre un emisor y un receptor cuyo principal objetivo es obtener el historial de actividad de los usuarios.

es solamente para esa empresa, en este caso Supercell, empresa finlandesa que ha obtenido los últimos años unos ingresos de ocho mil millones de dólares con los juegos Clash of Clans y Clash Royale (comentado anteriormente). Sin embargo, nuestros datos circulan por una serie de empresas que, aunque sean socias de Supercell nuestra información ya viaja por todo el mundo, si continuamos leyendo la política de privacidad, más abajo nos dice: “Supercell tiene socios que realizan servicios para nosotros. Estos socios procesan su información solamente bajo las instrucciones de Supercell y de acuerdo con estas para proporcionarle el Servicio, por ejemplo, el hosting, soporte técnico para jugadores, publicidad, analítica y prevención de fraude. Para combatir el fraude y la actividad ilegal, podremos intercambiar información con otras empresas y organizaciones y proporcionarla a autoridades en respuesta a solicitudes lícitas”.

Cabe destacar la compra de Supercell por Tencent. Esta última es la compañía china de videojuegos más grandes del mundo y creadora de productos como WeChat o QQ. Estos productos son las plataformas de comunicación más representativas y de mayor importancia de China. ¿Por qué esta compra? Desde el punto de vista de la compañía finlandesa está claro, aumentan su mercado de consumidores trasladándose a uno de los países más poblados del mundo, como dijo Ilkka Paananen⁹: "Esta nueva sociedad nos ofrece grandes oportunidades de crecimiento en China, donde vamos a poder alcanzar a cientos de millones de jugadores nuevos mediante los canales de Tencent". Sin embargo, la compañía china se dedicaba al sector de las comunicaciones y se ha extendido a la industria de los videojuegos. Entre otros intereses que podría tener la gigante empresa de comunicaciones china, y que son habituales en este tipo de adquisiciones y operaciones, se encuentra la recaudación de ingentes cantidades de datos de los usuarios (más de 100 millones de personas utilizan a diario las aplicaciones de Supercell).

⁹ Ilkka Paananen: es un magnate de los negocios, empresario y CEO (director ejecutivo) de Supercell y Sumea.

4.2. Nuestros datos en las empresas

Google

Cada aplicación aplica sus propios procedimientos, en el caso de Internet las empresas ofrecen sus motores de búsqueda, como Google (entre otros), gratuitamente, sin embargo, recogen todas aquellas páginas en las que nos metemos y, según unos patrones y procedimientos, lo analizan y extraen información sobre nosotros. Analizando la política de privacidad de Google, podemos darnos cuenta de que nos dice lo siguiente: “Al crear una cuenta de Google, nos proporcionas información personal que incluye tu nombre y una contraseña. También puedes añadir un número de teléfono o datos de pago a tu cuenta. Aunque no hayas iniciado sesión en una cuenta de Google, también puedes proporcionarnos información como, por ejemplo, una dirección de correo electrónico para recibir actualizaciones sobre nuestros servicios. También recogemos el contenido que creas, subes o recibes de otros usuarios cuando utilizas nuestros servicios. Entre estos datos se incluyen los correos electrónicos que escribes y recibes, las fotos y los vídeos que guardas, los documentos y las hojas de cálculo que creas y los comentarios que publicas en los vídeos de YouTube.” Es decir, recogen todo lo que escribimos y recibimos siempre que utilicemos el usuario de Google, toda la Huella Digital que vamos dejando al realizar búsquedas por medio de este buscador es almacenada y analizada por esta empresa.

En el caso de Google, Google Ads Settings¹⁰ elabora un perfil básico en función de tus búsquedas en Internet: tu género, tu edad aproximada y algunos de tus gustos. Además, también te puede dar un historial de ubicación en el caso de que tengas activada la localización del dispositivo, es decir, sabe en todo momento el lugar en el que te encuentras y toda esa información recopilada es almacenada en una base de datos. También puede saber qué dispositivos han realizado una búsqueda con tu cuenta y qué aplicaciones utilizan tu cuenta de correo electrónico como acceso.

¹⁰ Google Ads Settings: información que tiene Google de cada usuario y que utiliza para mostrar anuncios y hacer que sean más útiles para cada uno en función de sus gustos, su edad, y otras características.

Google usa el alcance de sus productos para recopilar información detallada sobre los comportamientos online y en el mundo real de todos sus usuarios y los que no lo son directamente, y los utiliza para dirigirse a ellos con publicidad pagada por terceros.

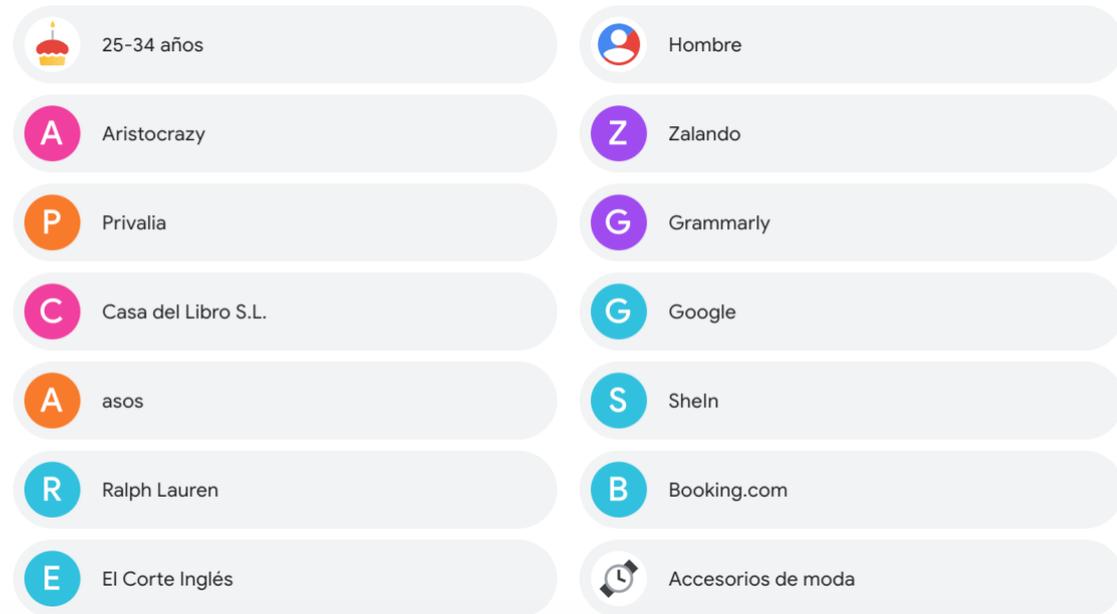


Ilustración 2 Perfil de un usuario de Google obtenido a través de Google Ads Settings

En esta imagen observamos el perfil que Google Ads Settings ha elaborado a través de determinados análisis de los datos. Este usuario tiene 26 años, es hombre y ha estado recientemente buscando información acerca de moda para hacer un regalo a una amiga. Además, ha estado buscando en diferentes páginas de hoteles para realizar un viaje al extranjero. Como podemos observar en la imagen, el análisis que ha realizado Google de este usuario a través de las búsquedas realizadas por este es bastante acertado. Todo esto ha hecho que a este usuario le aparezcan anuncios en relación con hoteles y moda (entre otros). Google se ha basado en la información personal y los intereses del usuario en cuestión y en los datos de los anunciantes de esta empresa para personalizar los anuncios que le resulten más útiles.

A continuación, veremos algunos de los motivos por los que nos aparecen anuncios en Google:

- Información de la cuenta con la que accedemos a Internet.
- Ubicación general.
- Consultas de búsqueda, tanto la actual como la anterior.
- Actividad con la sesión iniciada.
- Interacciones que haya tenido el usuario con los anuncios.
- Sitios web que son visitados.
- Actividad en aplicaciones en los diferentes dispositivos en los que tenemos abierta la cuenta.
- Hora del día.
- Información que entregamos a un anunciante cuando nos registramos con la dirección de correo electrónico.

El investigador Douglas C. Schmidt¹¹ ha publicado un estudio, “Google Data Collection”, en el que demuestra cuántos datos recopila cada dispositivo en un uso normal a través de cada uno de sus servicios. Google, además de ser propietario del primer navegador del mundo, proporciona la aplicación y plataforma móvil más instalada del mundo. Junto con los servicios de vídeo, correo electrónico y la aplicación de mapas, cuenta con más de mil millones de usuarios activos mensuales.

¹¹ Douglas C. Schmidt: científico informático y profesor en la Universidad de Vanderbilt. En “Google Data Collect” muestra la enorme cantidad de datos que esta empresa tiene de sus usuarios y sus hábitos más personales.

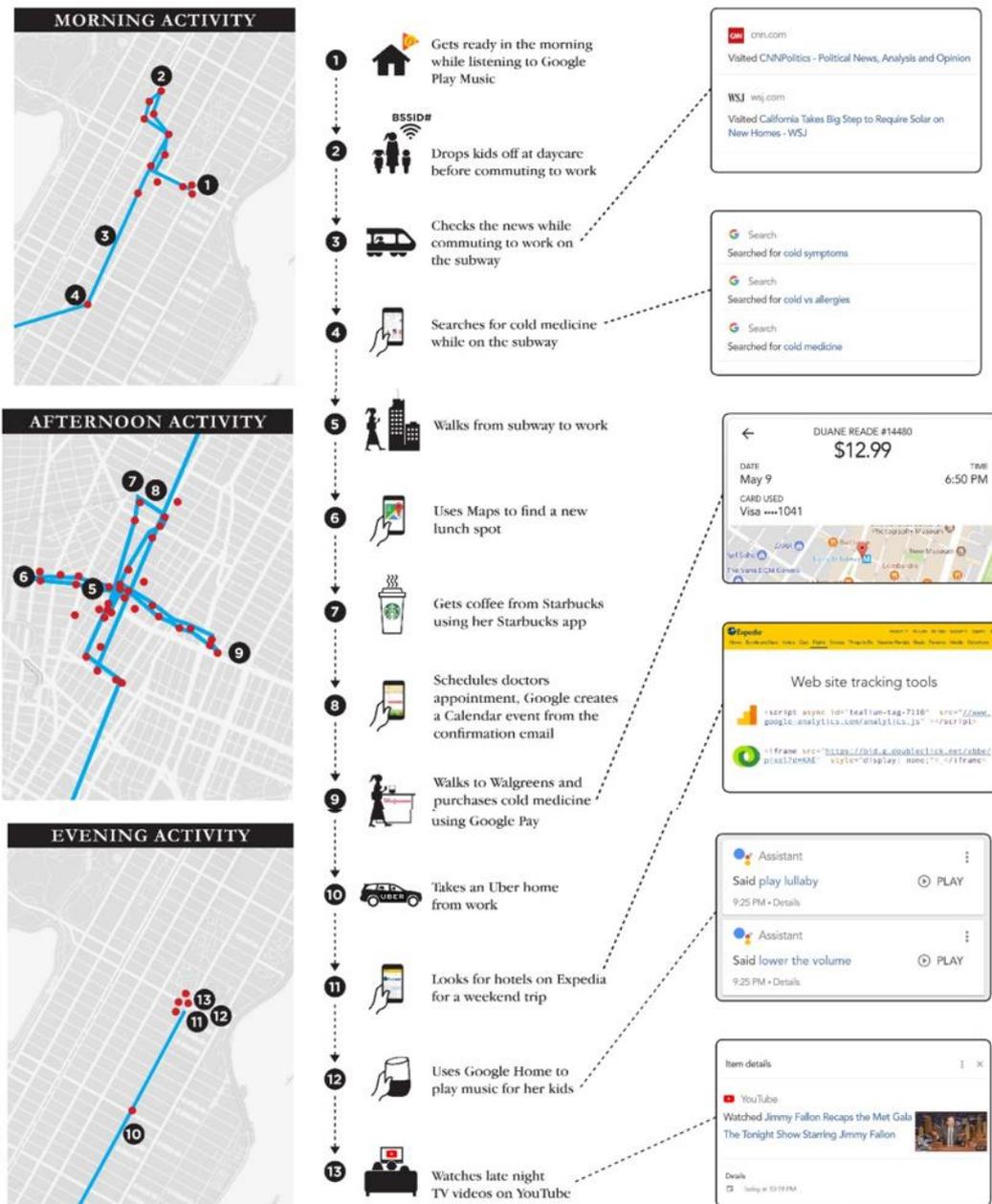


Ilustración 3 Análisis de los puntos clave de un usuario durante un día normal de la vida (Douglas. C. Schmidt, 2018)

Cuando nos creamos una cuenta de Google y habilitamos el historial de ubicaciones podemos encontrarnos con lo que aparece en la imagen superior. Si no lo deshabilitamos, seguirá todos nuestros pasos desde un móvil o cualquier otro dispositivo, ya que esta activado por defecto. Para poder ver realmente las fronteras a las que llegaba este seguimiento, he entrado en la web de un usuario (anónimo por motivos de privacidad) y

aparecía un mapa con puntos señalando los sitios en los que había estado. Otra de las cosas que me sorprendió fue que no solo registra cuando sales de una zona o una ciudad, sino que también registra los puntos en los que paras o los establecimientos por los que pasas (como vemos en la imagen superior). Todos estos datos recogidos son presentados en un orden cronológico con las horas de llegada y salida, si has ido andando, en coche o en otro medio de transporte. Otro aspecto sorprendente tiene que ver con una sección que se llama “My Activity”, el historial de esta sección recoge las búsquedas que hace cada uno en Google, Chrome o en un dispositivo Android. Es importante resaltar que no solo recoge las páginas en las que nos hemos metido, sino también la fecha y la hora.

Entre las condiciones de Servicio de Google nos encontramos con lo siguiente: “Al subir, almacenar o recibir contenido o al enviarlo a nuestros Servicios o a través de ellos, concedes a Google (y a sus colaboradores) una licencia mundial para usar, alojar, almacenar, reproducir, modificar, crear obras derivadas (por ejemplo, las que resulten de la traducción, la adaptación u otros cambios que realicemos para que tu contenido se adapte mejor a nuestros Servicios), comunicar, publicar, ejecutar o mostrar públicamente y distribuir dicho contenido. Google usará los derechos que le confiere esta licencia únicamente con el fin de proporcionar, promocionar y mejorar los Servicios y de desarrollar servicios nuevos”. Esto es, Google tiene derecho a utilizar nuestros datos libremente una vez que aceptamos sus términos y condiciones. Además de la información que consigue Google directamente de sus servicios y aplicaciones, recopila información de terceros. Un ejemplo de esto es que en 2014 anunció que comenzaría a rastrear las ventas en las tiendas físicas comprando datos de transacciones de tarjetas. Estos datos que contenían el nombre de la persona, la hora y ubicación de la compra, cubrieron el setenta por ciento de todas las transacciones en Estados Unidos.

Suena un poco desconcertante que una empresa, en este caso Google, tenga una cantidad ingente de información sobre nosotros. En 2016 la instalación de aplicaciones ascendía a 82.000 millones y se dio a conocer Google Play Protector, un nuevo sistema de seguridad que analiza las aplicaciones en segundo plano para comprobar que no se introduce ningún software malicioso; con el se escanean más de 1.000 millones de dispositivos al día, es decir, 50.000 millones de aplicaciones analizadas cada día. Google Photo llegó a la significativa cifra de 500 millones de usuarios activos en todo el mundo y que cada día

suben 1.200 millones de fotografías al servicio de almacenamiento de fotos. En 2017 las cifras de esta compañía alcanzaron los 2.000 millones de usuarios Android.

Facebook

Facebook es la empresa líder de las redes sociales que cuenta con más de 2.100 millones de usuarios en todo el mundo. Esta red permite contactar con personas conocidas de cualquier lugar e interactuar con ellas, compartir contenido, crear anuncios online, enviar mensajes y recibirlos, etc. tiene una gran variedad de funciones. Para una empresa puede resultar de gran utilidad ya que se puede utilizar para expandir un negocio o darse a conocer. Se podría decir que FB distingue entre el mundo de las empresas y el de las personas, pues ofrece dos posibilidades de perfil (perfil Personal y Fanpage). Entre las diferencias entre ambos destacan las siguientes:

- Perfil Personal: tipo de uso personal en el que los usuarios son amigos y la cantidad es limitada. Solo lo puede administrar una única persona y las opciones de personalización son básicas. Permite el envío de mensaje individual.
- Fanpage: tipo de uso profesional en el que los usuarios son seguidores y la cantidad de amigos es ilimitada. Puede ser administrado por múltiples usuarios, las opciones de personalización son avanzadas y permite el envío de mensajes en masa. Además, ofrece estadísticas en Facebook Insights, posee pestañas especiales y es posible crear anuncios a través de Facebook Ads.

Cuando nos creamos un perfil en FB proporcionamos nuestros datos personales como la cuenta de correo electrónico, nombre, edad y fotografía. Pero, además de estos datos, esta compañía también obtiene las conversaciones dentro de la aplicación, enlaces que compartimos, fotos que subimos y el lugar donde se ha hecho. Los sistemas de esta empresa analizan nuestra información, las páginas y las interacciones que tenemos dentro de esta aplicación, así como la frecuencia y duración de nuestras actividades. En el caso que usemos sus productos para realizar compras recopilarían y almacenarían esos datos de nuestras transacciones, y dispondrían de nuestros números de cuenta y tarjeta y detalles de la facturación y envío, por lo que también sabrían nuestra ubicación o, en el

caso de que no sea nuestra casa, ubicaciones de nuestro interés. No solo obtienen información nuestra directamente, si alguien nos menciona en un comentario o nos etiqueta en una imagen, obtienen igualmente información nuestra que es publicada por otras personas. Vivimos en un mundo interconectado en el que cada rastro de nuestra identidad es una huella digital que dejamos en la red. Atendiendo a la política de privacidad de FB vemos lo siguiente: “recopilamos información de los ordenadores, los teléfonos, las televisiones online y otros dispositivos conectados a internet que usas y que se integran con nuestros Productos, y combinamos esta información en los diferentes dispositivos que empleas”, es decir, tienen información de qué tipo de dispositivo estamos utilizando (Apple, Android, Tablet, ordenador o cualquier otro viable para esta aplicación), su sistema operativo, nivel de batería, aplicaciones que tenemos o potencia de la señal. Otro rasgo es que también saben si abrimos una ventana en segundo plano y cuánto tiempo o los movimientos que realizamos con el ratón.

Somos nosotros los que concedemos el permiso para que esta empresa disfrute de toda esta información. Somos libres de compartir nuestro contenido en la aplicación, si bien, FB debe contar con ciertos permisos de uso por parte de los usuarios. Cuando compartimos contenido protegido por los derechos de propiedad intelectual concedemos a esta empresa una licencia para poder usar, compartir y almacenar dicho contenido.

Facebook utiliza esa información recabada para enviar a sus usuarios mensajes de marketing y dar a conocer sus productos, así como para personalizar y mejorar estos de manera que proporciona a sus usuarios sugerencias de interés, su objetivo es crear productos personalizados para cada usuario. Los anuncios que se nos muestran guardan relación con noticias sobre acciones sociales que realizamos o que realizan nuestros amigos, como etiquetarnos en alguna pagina, enlace o fotografía, o darle a un “me gusta”. También deciden qué tipo de anuncios mostrarnos en función de la actividad en la web y aplicaciones ajenas a FB. Además, utiliza la actividad que tenemos fuera de sus productos para ayudar a medir la eficacia de los anuncios de los socios, lo que les permite conocer qué tipo de personas interactúan con sus páginas y quién utiliza sus servicios.

Por otro lado, fomenta la seguridad y la protección verificando cuentas y actividades en la aplicación. Mediante un análisis exhaustivo de la información de la que disponen consiguen detectar spam¹² y experiencias negativas e investigando infracciones de la política de privacidad. Por último, y no por ello menos importante, utiliza la información para llevar a cabo proyectos de investigación e innovación en los campos de la tecnología, salud y bienestar públicos. Cuando se produce un desastre natural, muchas veces las organizaciones de ayuda no saben donde se encuentran las personas afectadas; FB utiliza los mapas de desastre¹³. Esta información es de difícil obtención por los métodos convencionales de manera que, a medida que se comparte más información en Facebook los datos pueden proporcionar gran ayuda a las organizaciones humanitarias.



Ilustración 4 Configuración de anuncios de la aplicación Facebook

Esta imagen ha sido sacada de la configuración de anuncios mi perfil de FB, en ella podemos ver que esta empresa utiliza no solo datos que nosotros proporcionamos al registrarnos, sino también datos que los anunciantes y otros socios le proporcionan sobre nuestra actividad más allá de la aplicación.

¹² Spam: grandes cantidades de envíos publicitarios realizados en Internet y que se envía a un gran número de destinatarios con fines comerciales.

¹³ Mapas de desastre: ofrecen información sobre dónde se ubican las poblaciones, cómo se mueven y donde se están registrando durante un desastre natural; población, movimiento y control de seguridad.

En 2012 la mayor red social del planeta, Facebook, compró Instagram por valor de 1.000 millones de dólares. En 2014 alcanzó más de 130 millones de usuarios por segundo que dan ocho mil “likes” y mil comentarios de media. FB se dio cuenta del potencial que tenía la red social debido a las numerosas interacciones por parte de los usuarios. Cuando se realizó la compra de Instagram Zuckerberg (2012) dijo: "No planeamos hacerlo muchas más veces, si es que lo hacemos otra. Pero facilitar la mejor experiencia para compartir fotos es una de las razones por las que a tanta gente le gusta Facebook y sabíamos que valía la pena juntar a estas dos compañías”, y dos años más tarde, el 20 de febrero de 2014, Facebook adquiere WhatsApp por una suma de 19.000 millones de dólares. WhatsApp era y es el servicio de mensajería instantánea más popular en todo el mundo y cada día se intercambian unos 50.000 millones de mensajes a través de esta aplicación. Lo que propulsó a Mark Zuckerberg a realizar esta compra fueron varios aspectos, entre ellos están:

- WhatsApp suma un millón de usuarios al día y tiene 450 millones de usuarios activos al mes.
- El 72% de los usuarios utilizan sus servicios a diario.

Tras esta compra Jan Koum, CEO de WhatsApp, manifestó: “no cambia nada para los usuarios”, sin embargo, sí que ha cambiado, y de forma considerable. El momento en que FB adquirió a la empresa de mensajería instantánea los datos y la información de todos los usuarios de WhatsApp pasó a pertenecer a la red social más grande del mundo y, como hemos mencionado anteriormente, esta comparte dicha información con socios externos y con anunciantes (a los que ofrecen informes para indicar qué tipo de personas ven sus anuncios y qué resultados generan). Por lo tanto, todas las conversaciones que mantenemos por esta aplicación de mensajería como nuestro número de teléfono (el cual no tenía FB ya que no era requisito para poder crearte un perfil) están circulando por todo el mundo sin que nos demos cuenta.

Estadísticas de Facebook

Registra 500.000 nuevo usuarios cada día

El 68% de los americanos está en Facebook

El 79% de todos los adultos en el entorno online utilizan Facebook

El 76% de los usuarios de Facebook miran la plataforma cada día

El promedio de amigos es de 155

La mitad de os usuarios de Internet que no utilizan Facebook viven con alguien que sí lo usa

De ellos, el 24% mira fotos y publicaciones desde la cuenta de esa persona

Se estima que hay 270 millones de cuentas falsas

Hay 60 millones de páginas activas de negocios en Facebook

Facebook tiene 5 millones de anunciantes activos

Ilustración 5 Estadísticas de la empresa Facebook (Brandwatch, 2019)

Amazon

Amazon es la empresa de venta por Internet más grande del mundo. Cuenta con una base de unos 30 millones de clientes y hace su negocio obteniendo un porcentaje por la venta de los productos que ofrece. Todo su negocio se basa en la venta online, es decir, para poder comprar por Amazon es necesario estar registrado. Sabemos que al registrarnos estamos dando, entre otros, nuestro correo electrónico, nombre, datos de la cuenta bancaria con la que realizaremos los pagos, teléfono móvil, edad y ubicación. La mayoría de las personas, al no leerse toda la política de privacidad, no se dan cuenta de que proporciona mucha más información de lo que piensan. Esta empresa recoge información de manera automática (como en el caso de Facebook mencionado anteriormente) como puede ser: dirección IP que se utiliza cuando nos conectamos a Internet, información sobre el dispositivo utilizado (si es un ordenador, un teléfono móvil u otro dispositivo) y ubicación de este y utilización de aplicaciones. Este es otro ejemplo de todos los datos que poseen las empresas sobre nosotros.

La política de privacidad de Amazon dice: “Trabajamos con terceros como anunciantes, editores, redes sociales, motores de búsqueda, proveedores de publicidad y empresas de publicidad que trabajan por su cuenta, para mejorar la relevancia de los anuncios que ofrecemos. A la hora de proporcionarte anuncios basados en intereses, no relacionamos tus interacciones en sitios web que no sean propiedad de u operados por Amazon con datos personales, y no facilitamos ninguna información personal a los anunciantes ni a sitios web de terceros que muestren nuestra publicidad basada en intereses”, con esto nos quiere decir que toda la información recogida por Amazon no es vendida a terceros y no trafican con ella. De lo que no somos conscientes es de lo que viene después: “Los anunciantes, o las empresas de publicidad que ofrecen sus servicios a esos terceros, en ocasiones utilizan cookies en el proceso de provisión de contenidos, incluyendo anuncios, directamente en tu navegador o dispositivo y, cuando esto ocurre, puede que estos reciban automáticamente una dirección IP. También podrían utilizar cookies para medir la efectividad de sus anuncios, mostrarte contenido más relevante y prestarte servicios en nombre de Amazon”, normalmente no solemos leer las Cookies que aparecen al meternos en una página, simplemente damos clic en aceptar para poder seguir navegando sin tener una barra abajo o arriba de la pantalla que nos moleste. Ese es el problema, en el momento que damos a aceptar estamos entregando toda nuestra información a otra empresa de la que no sabemos de su existencia, con un simple clic. Entre los terceros que pueden utilizar cookies cuando utilizamos los servicios de Amazon están:

- AdForm → <https://site.adform.com/>
- Adobe → <https://www.adobe.com>
- AppNexus → <https://appnexus.com>
- bRealTime → <https://www.brealtime.com/>
- C1X Inc → <https://c1exchange.com/>
- Cloudflare → <https://www.cloudflare.com/>
- Cobalt → <https://www.cdkglobal.com/solutions/digital-marketing>

- Comet → <http://coxmt.com>
- ComScore → <https://www.comscore.com>
- Convertro → <https://convertro.com>
- District M → <https://districtm.net/en/>
- DoubleVerify → <https://www.doubleverify.com/>
- Evidon → <https://evidon.com>
- Facebook → <https://www.facebook.com/>
- Google → <https://www.google.com/>

De entre todos estos los más conocidos son Facebook y Google. Realizando un análisis y explorando los demás que hemos mencionado hay que destacar que la mayoría son empresas de marketing y software digital. Ya que Amazon es una de las empresas de comercio en Internet más grandes del mundo, estos terceros quieren ofrecer cookies por medio de esta para así llegar a un número elevado de personas que acepten las cookies y poder acceder a una mayor cantidad de datos.

Otra de las características y de las más recientes de esta compañía es el asistente virtual, Alexa. Cuando decimos su nombre activa el micrófono y empieza a funcionar. Este asistente se puede vincular a cualquier dispositivo tecnológico, un enchufe inteligente, el teléfono móvil, relojes de pared, etc. Con esto, Amazon dispondría una ingente cantidad de datos sobre la vida rutinaria de cada uno de los usuarios que tienen a Alexa. Además, es de gran relevancia la unión de Alexa y Cortana, asistente virtual de Microsoft. Las dos empresas son gigantes tecnológicos y tienen un gran número de usuarios cada una de ellas, con esta fusión cada una de las empresas podrá acceder a la información de los usuarios de la otra y así ofrecerles servicios y anuncios en función de sus intereses.

Es de importancia mencionar que los altavoces de Amazon, Amazon Echo, están escuchando en todo momento lo que pasa alrededor esperando recibir una orden.

El 27 de diciembre de 2016 se produjo un asesinato en una vivienda de Bentonville, Arkansas. Estos altavoces están en todo momento a la escucha de palabras clave que lo activen para utilizarlo, y así los audios que reconoce pasan por los servidores de reconocimiento de Amazon. La policía de esta ciudad pidió a esta empresa las grabaciones para poder resolver este crimen, sin embargo, Amazon se negó debido a los problemas de privacidad que esto ocasionaría. Si bien, la policía utilizó otros dispositivos conectados a Internet como un medidor de agua, el cual registró que se habían consumido 529 litros en dos horas. En el caso en que Amazon hubiera entregado las grabaciones, esto hubiera servido como precedente en que las fuerzas de seguridad de cada país podrían utilizar estos dispositivos como aparatos de escucha, lo que podría alarmar a las personas, ¿quién no se sentiría intimidado al saber que le están escuchando cada palabra o sonido que dice en cada momento? Las empresas del mundo tecnológico están más involucradas hoy en día debido a la cantidad de información que poseen de nosotros. Estas compañías estudian casos de peticiones de información por parte de las fuerzas de seguridad y esto nos demuestra la cantidad de dispositivos que tenemos alrededor que recopilan información personal y no nos damos cuenta.

Humana

Una de las áreas donde la tecnología y el análisis de datos masivos está teniendo gran relevancia es en el mundo de la salud. Podemos hablar tanto de la genómica¹⁴, biotecnología, los wearables o la realidad aumentada. La genómica personal es uno de los pilares más importantes de la medicina preventiva pues permite prever futuras enfermedades y elaborar el tratamiento más elaborado para cada paciente.

¹⁴ Genómica: se encarga del análisis de códigos ADN y ARN para saber más acerca de los genes, cómo funcionan y que impacto tienen. Se requieren cientos de petabytes para almacenar todos los genes de una persona (entre veinte mil y veinticinco mil) y toda la información analizada.

Este estudio del genoma gracias al poder del análisis de grandes cantidades de datos ha derivado hoy en día a diversas aplicaciones como son, entre otras:

- Empleo de modelos predictivos para diagnosticar pacientes de alto riesgo.
- Clasificación de subtipos de enfermedades para seleccionar tratamientos más dirigidos para cada paciente.

Podemos decir que el Big Data es de gran utilidad y de gran importancia en el mundo de la medicina, si bien, vamos a centrarnos en la huella digital que dejamos cada uno de nosotros en este. Los wearables son dispositivos digitales que llevamos todo el día encima e interactúan con nosotros (gafas, pulseras, relojes inteligentes o zapatillas con GPS). Cuando nos ponemos una pulsera al salir a correr para que lleve un registro de nuestro ritmo cardíaco, o de las calorías que quemamos, o incluso de nuestras pulsaciones para ver posteriormente la evolución entre el principio y el final, todos esos datos recabados durante un ejercicio son almacenados por la aplicación que utilizamos. Si nos metemos en un teléfono en la aplicación de salud (en este caso he elegido un dispositivo iPhone), podremos ver la gran variedad de información que saben de cada uno de nosotros y que realmente no pensamos en ello. Y no solo cuando hacemos algún tipo de ejercicio, también podemos observar la distancia que hemos andado, los pasos que hemos dado, los pisos que hemos subido, así como el grupo sanguíneo al que pertenecemos, si tomamos algún tipo de medicación, teléfono de los contactos de emergencia que hemos elegido en caso de que nos pasara algo, el peso, la altura, fecha de nacimiento o si eres o no donante de órganos. En las imágenes que veremos a continuación nos damos cuenta de que esta aplicación recoge toda la información mencionada anteriormente, además de la hora de irse a la cama y la de levantarse mediante la actividad registrada en el teléfono.



Ilustración 6 Datos de salud y datos médicos de la aplicación Salud de un usuario de iPhone



Ilustración 7 Actividad registrada durante el día y análisis de sueño de un usuario de iPhone

Humana, fundada en 1961 en Louisville, Kentucky, es una empresa que comercializa y administra seguros de salud. Tiene un conjunto de 11,5 millones de clientes en Estados Unidos y comercializa sus seguros en los cincuenta estados. Habiendo visto todo esto, es de considerar la situación de Walter y Paula Shelton, una pareja de sexagenarios que vivía en Luisiana que estaba buscando empresas aseguradoras para poder tener un seguro médico. Esta pareja realizaba compras de medicamentos en grandes almacenes como WalMart y Randalls. La compañía de seguros Humana, donde se querían asegurar Walter y Paula, tras realizar un análisis del historial de recetas médicas decidió realizarles una serie de preguntas en función a unos medicamentos determinados, antidepresivos y medicación para la presión arterial. El marido explicó que los primeros los compraba debido a que Paula estaba con la menopausia y pudiera dormirse fácilmente. En cuanto a los segundos, explicó que se debía a una hinchazón de los tobillos. Finalmente, Humana negó a esta pareja el poder tener un seguro médico en su compañía. Esto es un claro ejemplo de la huella digital y el Big Data. La pareja de sexagenarios fue dejando un rastro durante años sobre los productos que compraba en los almacenes mencionados anteriormente y, para la empresa, el análisis de datos ha facilitado la toma de una decisión. Según Bloomberg¹⁵ dos tercios de las empresas de seguros estadounidenses utilizan las bases de datos de prescripciones médicas para analizar a sus futuros clientes y tomar decisiones basándose en ellas.

¹⁵ Bloomberg: fundada por Michael Bloomberg en 1981, es una empresa que se dedica a ofrecer servicios de información financiera. Permite consultar y analizar información financiera en tiempo real y realizar operaciones de compra venta de activos a través de su plataforma electrónica.

5. Como afectan a las empresas las leyes de protección de datos

La Ley Orgánica 15/1999 fue aprobada por las Cortes Generales el 13 de diciembre de 1999 y tenía por objetivo garantizar y proteger el tratamiento de los datos personales, los derechos fundamentales y las libertades públicas y sobre todo de su intimidad y privacidad personal y familiar. Esta ley se fundamentó en el artículo 18 de la constitución española de 1978 (derecho a la intimidad familiar y personal y e secreto de las comunidades). Actualmente vivimos en un mundo tecnológico y sujeto al cambio constantemente, a diario se publican comentarios, mensajes e imágenes que son almacenados en bases de datos y analizados posteriormente por las empresas. Ante este avance tecnológico y el afán de obtener cantidades ingentes de datos por parte de las empresas se elaboró la Ley Orgánica de Protección de Datos y de Garantía de Derechos Digitales que entró en vigor el 6 de diciembre de 2018. Esta ley regula por primera vez los derechos digitales y establece nuevos derechos para los trabajadores como la desconexión digital. Un ejemplo de esto es la empresa Telepizza, en la que sus repartidores estaban obligados a utilizar su teléfono móvil personal para tener un seguimiento de los pedidos por parte de la empresa y de los clientes, así mismo esta empresa pagaba una suma de tres con cincuenta euros por el consumo de datos del teléfono. La finalidad de esto es que los clientes pudieran saber en tiempo real el estado de su pedido, dentro del contrato estaba establecido la descarga obligatoria de esta aplicación que permite rastrear el dispositivo personal de cada uno de los repartidores y en caso de negarse se procedía al despido. Por ello, la LOPDGDD sostiene que los trabajadores tienen derecho a la desconexión digital para respetar la intimidad personal y familiar de los trabajadores. Mediante este derecho se conseguirá una mejor calidad de vida del empleado, aumento de a productividad, reduce la rotación de personal y el absentismo laboral y una mayor incentivación de la conciliación laboral de los trabajadores. Además, la LOPDGDD permite a la empresa acceder al contenido de los dispositivos solo y exclusivamente con la finalidad de controlar el cumplimiento de las obligaciones laborales.

El artículo 90 de la Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales especifica la información que se debe entregar a los trabajadores antes de implantar un dispositivo de geolocalización en su dispositivo y es la siguiente:

1. “Los empleadores podrán tratar los datos obtenidos a través de sistemas de geolocalización para el ejercicio de las funciones de control de los trabajadores o los empleados públicos previstas, respectivamente, en el artículo 20.3 del Estatuto de los Trabajadores y en la legislación de función pública, siempre que estas funciones se ejerzan dentro de su marco legal y con los límites inherentes al mismo”.
2. “Con carácter previo, los empleadores habrán de informar de forma expresa, clara e inequívoca a los trabajadores o los empleados públicos y, en su caso, a sus representantes, acerca de la existencia y características de estos dispositivos. Igualmente deberán informarles acerca del posible ejercicio de los derechos de acceso, rectificación, limitación del tratamiento y supresión”.

Actualmente la información es uno de los principales activos de una empresa, es por esto por lo que quieren obtener el máximo de datos posibles tanto de sus trabajadores como de sus clientes, para así poder establecer una mejor segmentación de sus clientes, reclutar empleados de mayor calidad y capacidad y mejorar la toma de decisiones. Sin embargo, la aparición de este tipo de leyes y normas, como el Reglamento General de Protección de Datos, va a dificultar la obtención de información por parte de las empresas y va a proteger más la privacidad de las personas.

6. Resumen y conclusiones

El objetivo de este trabajo ha sido explicar qué es el Big Data y la huella digital, así como la influencia de los datos en la sociedad. Durante la realización de este trabajo se ha explicado el concepto de BD y las características principales de los datos (volumen, variedad, velocidad y variedad). Además, se ha realizado un estudio cronológico desde la aparición del BD hasta hoy. Hemos visto la relevancia que tienen los datos en las empresas y se han identificado diversas aplicaciones que pueden ser utilizadas en el mundo actual; conocimiento de clientes, recursos humanos, medicina, optimización de máquinas y de procesos estratégicos y en el mundo del deporte. Visto esto, es indudable que el análisis de datos masivo está presente en casi todos los sectores. En base a la importancia de los datos en el mundo en el que vivimos se ha considerado de gran relevancia abordar el tema del Big Data y de la Huella Digital.

Al utilizar un dispositivo tecnológico dejamos un rastro, simplemente con hacer clic en un enlace o una aplicación, ya estamos transmitiendo información a una base de datos que luego utilizarán para las empresas para orientar contenidos y productos hacia un mercado y consumidores específicos. Es desconcertante la cantidad de datos que tienen terceros sobre nosotros y que se desplazan por todo el mundo. Entre las vías más fáciles de captación de datos destacan: las redes sociales, ya que registran todos tus datos necesarios para poder acceder, además de los comentarios y las fotografías subidas con las que pueden saber en todo momento el lugar donde te encuentras; y las búsquedas en Internet, los buscadores almacenan todas las búsquedas realizadas y mediante algoritmos y patrones analizan los intereses y gustos de las personas determinando así qué tipo de anuncios mostrarle. La ausencia de riesgo se muestra en nuestro comportamiento en Internet. Pensamos que cuando compartimos nuestros datos en una red social o un servicio el problema está en configurar las preferencias de privacidad que el prestador nos proporciona. Cada vez somos menos conscientes de cuánto de nuestra vida exponemos y, por lo tanto, mantener fuera del alcance de otros nuestros datos personales es cada vez más complicado. Siempre escuchamos que es peligroso subir todo tipo de contenido a las redes sociales y que hay que regularlo, pero seguimos sin hacer caso. Cuando damos nuestros datos estamos confiando en una empresa de la que, probablemente, no sabemos nada; no sabemos si es seguro que tengan nuestra

información. En cierto modo es porque no llegamos a percibir ese riesgo como algo cercano y que nos pueda pasar, por el contrario, pensamos que es ajeno a nosotros.

Hoy en día somos visibles en casi todas las interacciones con plataformas tecnológicas. Somos rastreados y analizados en todo momento por las grandes oportunidades que tiene acceder a una cantidad enorme de información para las empresas. Las nuevas leyes de protección de datos y de privacidad de las personas suponen un gran muro entre la obtención masiva de datos y las empresas, de entre ellas el Reglamento General de Protección de Datos es la norma de más relevancia en el ámbito de la privacidad.

7. Bibliografía

Agencia Española de Protección de Datos, AEPD. (2019): “*Fingerprinting o Huella digital del dispositivo*”. Madrid.

Alonso, Óscar. (2018): “*Monetización de los datos: la importancia de los datos para las empresas*”. [Documento de Internet disponible en <https://www.lavanguardia.com/economia/20180104/434058696189/monetizacion-datos-empresas-the-valley.html>].

Caballero, Rafael. Martín, Rafael. (2015): *Las bases de Big Data*. Madrid: La Catarata.

Calonge, Iciar. (2015): “*Business Intelligence y Big Data. ¿Son lo mismo?*”. [Documento de Internet disponible en <https://www.conasa.es/blog/business-intelligence-y-big-data-son-lo-mismo/>].

Curto Díaz, Josep. (2010): *Introducción al Busines Intelligence*. Barcelona: UOC (Universitat Oberta de Catalunya).

Duhigg, Charles. (2012): “*How Companies Learn Your Secrets*”. [Documento de Internet disponible en <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>].

Galdon, Gema. (2015): “¿Qué hacen con nuestros datos en internet?”. [Documento de Internet disponible en https://elpais.com/tecnologia/2015/06/12/actualidad/1434103095_932305.html].

Goizueta, Javier. (2018): “*El RGPD pone en jaque el gran negocio del siglo XXI*”. [Documento de Internet disponible en https://blogs.elconfidencial.com/espana/tribuna/2018-09-13/datos-rgpd-gran-negocio-sigloxxi_1614624/].

González, Alberto. (2019): “¿Es legal la geolocalización de los trabajadores?”. [Documento de Internet disponible en <https://elderecho.com/legal-la-geolocalizacion-los-trabajadores/>].

Hierro Álvarez, Jorge. (2016): “Cómo definir una estrategia a través del Big Data en la empresa”. [Documento de Internet disponible en <https://www.iebschool.com/blog/estrategia-a-traves-del-big-data/>].

Joyanes Aguilar, Luís. (2013): *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. México: Alfaomega Gripo Editor, S.A. de C.V.

Lohr, Steve. (2016): “Más y mejor seguridad para poder vivir en un mundo del internet de las cosas”. [Documento de Internet disponible en <https://www.nytimes.com/es/2016/10/22/mas-y-mejor-seguridad-para-poder-vivir-en-un-mundo-del-internet-de-las-cosas/>].

Luna, Juan José. (2018): “Las tecnologías Big Data” [Documento de Internet disponible en <https://www.teldat.com/blog/es/procesado-de-big-data-base-de-datos-de-big-data-clusters-nosql-mapreduce/>].

Marr, Bernard, (2016): *Big Data: La utilización del Big Data, el análisis y los parámetros decisiones SMART para tomar mejores y aumentar el rendimiento*. Zaragoza: S.L. TEELL EDITORIAL.

Marr, Bernard, (2017): *Big data en la práctica: Cómo 45 empresas exitosas han utilizado análisis de big data para ofrecer resultados extraordinarios*. Zaragoza: S.L. TEELL EDITORIAL.

Martínez, Jesús. (2017): “Intsagram: cinco años de la compra más rentable de Zuckerberg en Facebook”. [Documento de Internet disponible en https://www.elespanol.com/economia/empresas/20170331/204980067_0.html].

Mayer-Schonberguer, Viktor. Cukier, Kenneth. (2013): *Big data: La revolución de datos masivos*. Madrid. Turner Publicaciones S.L.

Monleón-Getino, Antonio. (2015): *El impacto del Big-data en la Sociedad de la Información. Significado y utilidad*. Barcelona: Universidad de Barcelona.

Moorthy, Janakiraman. Lahiri, Rangin. Biswas, Neelanjan. Sanyal, Dipyaman. Ranjan, Jayanthi. Nanath, Krishnadas. (2015): Big Data: Prospects and Challenges. *The Journal for Decision Makers*,(40), 74-96.

Pasquale, Frank. (2015): *THE BLACK BOX SOCIETY The Secret Algorithms That Control Money and Information*. Londres: Harvard University Press. .

Pole, Andrew. (2007): *Statistical Arbitrage: Algorithmic Trading Insights and Techniques*. Estados Unidos: John Wiley & Sons, Inc.

Rochina, Paula. (2016): “*Nuestra huella digital en Internet: ¿Hasta dónde saben de mí?*”. [Documento de Internet disponible en <https://revistadigital.inesem.es/informatica-y-tics/huella-digital-internet/>].

Rodríguez, Sergio. (2012): “*Facebook compra Instagram por 1.000 millones de dólares*”. [Documento de Internet disponible en <https://www.elmundo.es/elmundo/2012/04/09/navegante/1333991473.html>].

Rojas, Elisabeth. (2012): “*Las 3 uves del Big Data*”. [Documento de Internet disponible en <https://www.muycouterpro.com/2012/11/08/uves-big-data>].

Rubio, Isabel. (2019): “*Empleados de Amazon escuchan a diario conversaciones que mantienen los usuarios con Alexa*”. [Documento de Internet disponible en https://elpais.com/tecnologia/2019/04/11/actualidad/1554992401_521050.html].

S. Mazo, Estela. (2014): “*Facebook compra WhatsApp por 19.000 millones de dólares*”. [Documento de Internet disponible en <http://www.expansion.com/2014/02/19/empresas/tmt/1392849185.html>].

Sáez, Cristina. (2012): “¿Dónde están mis datos?”. [Documento de Internet disponible en <https://www.lavanguardia.com/estilos-de-vida/20120719/54326650737/donde-estan-mis-datos.html>].

Sánchez, Analía. (2013): “¿Por qué es importante Big Data”. [Documento de Internet disponible en <https://dataiq.com.ar/blog/por-que-es-importante-big-data/>].