GRADO EN COMERCIO

TRABAJO FIN DE GRADO

Big Data y su influencia en el consumo en la empresa privada

CRISTINA RIVAS GONZÁLEZ

VALLADOLID, 2019

FACULTAD DE COMERCIO
Universidad de Valladolid

UNIVERSIDAD DE VALLADOLID
GRADO EN COMERCIO

CURSO ACADÉMICO 2018/2019

TRABAJO FIN DE GRADO

Big Data y su influencia en el consumo en la empresa privada

Trabajo presentado por:

CRISTINA RIVAS GONZÁLEZ

Firma:

Tutor:

OSCAR M. GONZÁLEZ RODRÍGUEZ

Firma:

Valladolid, 2019

FACULTAD DE COMERCIO
Universidad de Valladolid

Índice

Índice de figuras

1 INTRODUCCIÓN

Vivimos en una era digital, es un hecho. Pasamos varias horas al día navegando en Internet con nuestro Smartphone, y al igual que cada movimiento que hacemos, todo deja huella. En este caso es una huella digital, pero el Big Data va más allá.

Durante los últimos años, la palabra “Big Data” ha aparecido en muchos contextos, en el periódico, en la televisión, incluso en nuestro día a día y cada vez más en las universidades. Sin embargo, al igual que muchas personas, aunque había oído hablar del Big Data, no me quedaban claras sus funciones ni aplicaciones en la empresa.

En el Grado de Comercio nos proporcionan conocimientos de las diferentes áreas de una empresa como contabilidad, marketing, dirección de empresas o toma de decisiones… sin embargo, a pesar de conocer todos estos ámbitos, siento que la empresa avanza muy rápido y en muchos casos necesitan adaptarse. Tras indagar un poco en las diferentes fuentes de información y en el análisis de datos para beneficio de éstas, me topé repetidas veces con el concepto “Big Data”, es decir, procesamiento de información de forma masiva. Me di cuenta de que todas nuestras conductas se convertían en datos, con un enorme valor, y que las empresas cada vez se beneficiaban más de ellos para mejorar su negocio o tener una ventaja competitiva. Sin embargo, el Big Data no se queda aquí, no solo las empresas se benefician de estos sistemas, sino que ya se utiliza en otros ámbitos como los deportes o en elecciones presidenciales.

Al encontrarme con toda esta información, tuve muy claro que mi Trabajo de Fin de Grado tenía que ir orientado a incrementar los conocimientos obtenidos durante los cuatro años de carrera, y el Big Data es uno de los pilares fundamentales de toda gran empresa, sea cual sea su negocio. Por ello, encuentro muy interesante el poder conocer cómo aplicar estas tecnologías en los negocios y qué beneficios nos puede aportar, para entender mejor por qué a día de hoy, el valor de la información es una prioridad para las empresas.

He querido desarrollar mi trabajo de manera muy intuitiva y escalonada, comenzando con una presentación de conceptos y del contexto actual. El grado de Comercio no es una carrera científica “pura”, por lo que no quería orientar el trabajo a datos técnicos de funcionamiento de las aplicaciones o proceso Big Data, sino a conseguir un conocimiento general del concepto. En base a ello, he desarrollado posteriormente un trabajo de campo donde busco un punto en común entre los conocimientos sobre Big Data de la población en Valladolid, y la opinión de profesionales que trabajan en su día a día con estos sistemas.

Objetivos

Antes de comenzar a indagar y buscar información sobre el Big Data, había muchas preguntas que venían a mi mente. ¿Qué se necesita para implantar sistemas Big Data en la empresa? ¿de dónde obtienen información? ¿por qué la información es tan valiosa para las empresas? ¿Qué beneficio obtenemos nosotros como consumidores? ¿todas las empresas deberían implantar sistemas Big Data? ¿A partir de qué nivel de datos consideramos que es necesario utilizar Big Data? ¿Hay otras alternativas? ¿Quiénes se benefician de ello?

El objetivo principal de mi Trabajo de Fin de Grado es demostrar y encontrar un sentido común a la realidad que existe en las empresas españolas en relación a sus conocimientos y consideración de la implantación de sistemas Big Data, con la realidad del conocimiento que tiene la población y cómo influye en el consumo. Para conseguirlo, me propuse una serie de pequeños objetivos en base a los cuales se ha desarrollado el trabajo:

- Entender el concepto Big Data y sus fases, desde la obtención de la información hasta la obtención de conclusiones que nos ayudan posteriormente a la toma de decisiones.
- Conocer las principales aplicaciones y tecnologías utilizadas a lo largo de todas las fases, así como su funcionamiento.
- Analizar y entender los beneficios e inconvenientes que proporciona a las empresas la implantación de sistemas Big Data.
- Conocer la situación actual del Big Data en España, en comparación con otros países europeos y EEUU.
- Comprender las limitaciones que proporciona la actual ley de protección de datos a las empresas que comercializan con información y los límites de éstas.
- Investigar a través de una encuesta la opinión y los conocimientos de los habitantes de Valladolid sobre la utilidad, el funcionamiento y la realidad actual del Big Data.
- Estudiar la opinión de expertos sobre la realidad de la implantación de sistemas de análisis Big Data en las empresas y su tendencia de desarrollo actual.

Agradecimientos

Me gustaría agradecer a varias personas la ayuda y el apoyo que me han prestado durante estos meses de trabajo.

Principalmente a mi tutor Oscar M. González Rodríguez, por aceptar el desafío de tutorizarme a distancia, por su apoyo y constancia.

Agradecer también a mis padres, que me han motivado siempre a crecer y aprender.

Y a mis amigos, por creer siempre en mí.

2 BIG DATA: CONCECTOS BÁSICOS

El siglo XXI es la era de las nuevas tecnologías, la economía 4.0 o "cuarta revolución industrial" con su principal cambio en la transformación digital. La evolución actual hacia una era digital, ha traído consigo cambios importantes dentro de la empresa como la modificación de las estrategias empresariales, que van ligadas a la introducción de la robotización en la empresa y la automatización de la producción, para conseguir una producción más barata y eficiente.

Estas modificaciones a la hora de tomar decisiones van de la mano de la implantación de un nuevo sistema o tecnologías basadas en el procesamiento de datos de forma masiva, el Big Data. A través de las nuevas conductas que llevamos a cabo, véase el uso de redes sociales y de Internet en nuestro día a día, la monitorización de nuestras decisiones y todos los dispositivos electrónicos que hemos introducido a nuestra vida, han generado una cantidad tan grande de información que las técnicas habituales de análisis se han quedado obsoletas. Por ello, es necesario utilizar otro tipo de herramientas y fórmulas adaptadas a la nueva economía, para conseguir un buen procesamiento de datos que posteriormente se convierta en una información valiosa para la empresa. Esto nos deja varias preguntas, la primera concierna a las dificultades de implementación hoy en día, las áreas donde es aplicada y sus posibles beneficios. Así como cuál es la manera más efectiva de poder visualizar y estudiar esta información para conseguir un mayor aprovechamiento del contenido.

Han sido las grandes empresas las primeras en adoptar y tener presentes en sus previsiones un presupuesto para el procesamiento de datos (Big Data), utilizando su posición dominante y de poder en el mercado para acceder más fácilmente a ellos. De hecho, las cinco empresas más grandes del mundo hoy en día procesan datos de forma masiva. Y aunque tienen diferentes fuentes de ingresos, empresas como Apple, Microsoft, Amazon, Facebook y Google estudian principalmente patrones de comportamientos sociales y lo convierten en valor económico. Sin embargo, si hablamos de empresas más pequeñas, tanto las pymes como el sector público han tenido que adaptar su negocio implantando algunas de estas tecnologías, pero sobre todo aprendiendo y probando el manejo de nuevas técnicas necesarias para sobrevivir a esta revolución tecnológica. Esto tiene como objetivo ayudar a las empresas a crecer y a proporcionarles una información valiosa y necesaria para una mejor toma de decisiones, una gran ventaja sobre las empresas que no lo hacen, y un gran poder, para anticiparse y tomar decisiones más rápidas, objetivas y rentables.

A priori se puede presentar como una solución con grandes ventajas, pero como todo, tiene un lado negativo, como, por ejemplo, la rapidez con la que se está desarrollando conlleva una falta de adaptación de las leyes que lo regula, lo que puede generar muchos problemas y desconfianza en la sociedad. La necesidad de establecer unos límites tanto legales como morales dentro del Big Data se encuentra entre las prioridades de estos nuevos cambios, y uno de los principales actores y responsables son las empresas privadas.

2.1 Introducción al Big Data

Todos los expertos coinciden en que el Big data nació con Google en 2003 para cubrir unas necesidades que no estaban cubiertas por la existente tecnología del momento, como era el almacenamiento y análisis de grandes cantidades de datos, y que poseían unas características muy particulares. El Big Data supone la confluencia de numerosas tendencias tecnológicas que se llevaban varios años consolidando en nuestra sociedad y que han irrumpido de manera muy fuerte en las organizaciones y empresas, tanto públicas como del sector privado, pero sobretodo han irrumpido en la sociedad a través de las redes sociales, la reducción del coste de acceso a Internet, el Internet de las cosas, la geolocalización y de manera muy significativa la computación en la nube **(Debitoor, 2019).**

Hoy en día, los volúmenes de datos han ido creciendo de manera espectacular. Antes de hablar sobre la cantidad de datos creados durante los últimos años y los que se crearán en un futuro, es importante comprender las medidas de éstos. Al hablar de Big Data, lo analizaremos en términos de bytes:

Gigabyte = 10^9 = 1,000,000,000 bytes
Terabyte = 10^{12} = 1,000,000,000,000 bytes
Petabyte = 10^{15} = 1,000,000,000,000,000 bytes
Exabyte = 10^{18} = 1,000,000,000,000,000,000 bytes

Para hacernos una idea de las grandes cantidades de las que estamos hablando, durante el 2012 se crearon 2,8 zettabytes (ZB) de datos, lo cual corresponde a 2,8 millones de gigabytes según los datos de la consultora IDC en el estudio “El universo digital de datos 2012” y esta cifra se dobla cada dos años **(Guilarte, 2012).** Otro dato de importancia que continúa ilustrando la fuerza y rapidez con la que se crean datos, es Walmart, una gran cadena de almacenes de Estados Unidos, que según Luis Joyanes en su libro “Big Data, Análisis de grandes volúmenes de datos en organizaciones”, posee bases de datos con una capacidad de 2,5 petabytes y lleva a cabo más de un millón de transacciones cada hora **(Joyanes 2013).**

Es en el año 2000 cuando el analista de la industria Doug Laney[1] proporcionó una definición muy particular basada en las 3 Vs **(SAS, 2019)**:

- **Volumen:** muchas veces es utilizado como sinónimo de Big Data. Hace referencia al tamaño de los datos (terabytes, exabytes, petabytes), creado tanto por personas como por máquinas. Estos provienen de diversas fuentes, y aunque almacenarlos en el pasado hubiera supuesto un gran problema, actualmente es bastante sencillo gracias a tecnologías como Hadoop[2]. La cantidad de datos que se genera actualmente es tan grande que era inimaginable hace algunos años. Estamos conectados en el mundo 2.0, donde para muchas empresas estar presentes en el mundo digital es algo imprescindible, por ejemplo, para las empresas que únicamente venden a través del canal online, necesitarán implantar sistemas de gestión de datos Big Data para poder recoger todos los datos que se genera cada día en su página **(iic, 2016).**
- **Velocidad**: se asocia con la rapidez con la que se generan y llegan los datos. Cada día es posible la obtención de datos con una mayor rapidez, es el caso de los chips GPS y el alcance de las redes inalámbricas, con lo que es posible conocer la posición de un vehículo al momento con relativa precisión y almacenarla fácilmente **(Galimany, 2014).**
- **Variedad**: va muy ligado con el volumen, explica la variedad de datos que se genera, los cuales vienen en toda clase de formatos, podemos hablar de:
    - **Estructurados**: son aquellos datos que contienen un formato o esquema fijo. Las mayorías de las fuentes de datos tradicionales son datos estructurales. Poseen un formato bien definido y un orden específico. Los formatos típicos que se encuentran dentro de este tipo de datos son las fechas de nacimiento (DD, MM, AA), el número de cuenta corriente (20 dígitos) o el documento de identidad (8 dígitos y una letra).
    - **Semi-estructurados**: tienen un flujo lógico y un formato definido pero cuya comprensión no es fácil. No tienen formatos fijos, pero contienen etiquetas o algunos marcadores que nos permite definir su formato y separar sus elementos. Dentro de este tipo de datos encontramos el texto de etiquetas de lenguajes XML y HTML.

[1]Investigador vicepresidente y distinguido analista en la empresa Gartner

[2]FrameWork que permite el procesamiento de grandes volúmenes de datos a través de grupos de ordenadores que utilizan modelos de programación simple.

- **No estructurados**: son aquellos datos que no siguen un tipo predefinido. Se almacenan sin ninguna estructura uniforme y se tiene poco control sobre ellos. Aquí es donde se encuentra la mayor parte de los datos que se generan y los que poseen las organizaciones, sin duda, son los más difíciles de manejar, y es a partir de su aparición cuando el Big Data ha comenzado a tomar importancia, necesitando de herramientas para su manipulación e interpretación. Algunas de estas herramientas desarrolladas son MapReduce, Hadoop o bases de datos NoSQL. Algunos ejemplos de fuentes de donde se producen estos datos son los mensajes de texto, los sms, la mensajería instantánea tipo WhatsApp, los audios, videos y fotografías.

Para entender mejor los diferentes tipos de datos que podemos encontrar, vamos a ver el caso de la compañía de Facebook. Es fácil imaginar la gran cantidad de información y la diversidad de ésta viendo la cantidad de información que cada uno de nosotros puede generar al día y viendo los millones de usuarios diarios que utilizan la red social. En su base de datos podemos encontrar datos estructurados como son los presentes en los sistemas corporativos (fecha de nacimiento, edad…), datos semi-estructurados que se encuentran en los registros del sistema como servidores web y los datos no estructurados que tienen un formato de imagen, audio o video entre otros muchos.

Al principio, esta capacidad de análisis se limitaba tan solo a los datos estructurados y almacenados en base de datos relacionales (basados en el uso de tablas, registros y columnas). Con el Big Data y sus nuevas herramientas asociadas, se ha introducido la capacidad de combinar datos de orígenes distintos y de formatos no homogéneos, lo que ha permitido aumentar la capacidad de análisis.

Estas características pueden quedarse un poco ambiguas, y actualmente hablamos más de 5Vs, que además nos permite realizar una definición más completa sobre el Big data. A las tres características explicadas anteriormente añade las dos siguientes **(Joyanes 2013):**

- **Veracidad**: tenemos la certeza de que los datos que estamos analizando son reales. Según qué tipos de datos, se le puede aplicar algún método para "limpiarlos", pero a veces es imposible ya que no se puede eliminar la imprevisibilidad que existe en variables como el comportamiento futuro en una determinada parte de la población, o la situación política en un país.
- **Valor**: estos datos poseen mucho valor, tanto en términos monetarios como informativos. Es el propósito final del Big Data, ya que los datos sin

combinarlos, transformarlos y analizarlos no poseerían ningún valor. Lo que las empresas buscan a través del Big Data es utilizar estos datos para obtener información de forma fiable. Esto puede llegar a suponer un problema, ya que las empresas deben buscar la forma más coherente y alienada con la estrategia empresarial de la misma para poder sacar el máximo beneficio y provecho a estos datos.

2.2 ¿Qué es el Big Data?

Después de realizar una pequeña introducción al Big Data, podemos incluir todas estas características en una sola definición. El IBM (International Business Machines Corporation) nos propone la siguiente reflexión de la cual se extrapola la definición: "Además del gran ***volumen*** de información, esta existe en una gran ***variedad*** de datos que pueden ser representados de diversas maneras en todo el mundo, por ejemplo de dispositivos móviles, audio, video, sistemas GPS, incontables sensores digitales en equipos industriales, automóviles, medidores eléctricos, veletas, anemómetros, etc., los cuales pueden medir y comunicar el posicionamiento, movimiento, vibración, temperatura, humedad y hasta los cambios químicos que sufre el aire, de tal forma que las aplicaciones que analizan estos datos requieren que la ***velocidad*** de respuesta sea lo demasiado rápida para lograr obtener la información correcta en el momento preciso" **(IBM, 2012)**.

Existen muchas otras definiciones sobre el Big Data, como la proporcionada por el vicepresidente de la consultora Gartner, Adrián Merv, **(Gartner, 2012):** "Son activos de información caracterizados por su alto volumen, velocidad y variedad, que demandan soluciones innovadoras y eficientes de procesado para la mejora del conocimiento y toma de decisiones en las organizaciones."

El IDC[3] los define como "la capacidad de analizar y tratar un gran volumen de datos con el objetivo de convertirlos en valor para la toma de decisiones" **(IDC, 2018).**

En CIO[4] Hopkins define Big Data como "las técnicas y tecnologías que hacen que sea económico hacer frente a los datos a una escala extrema.

[3] International Data Corporation. Primer proveedor de inteligencia de mercado, servicios de asesoría y eventos para los mercados de tecnología de la información, telecomunicación y tecnología de consumo.

[4] Revista líder de gestión estratégica de la tecnología.

Big Data trata de tres cosas:

1) Las técnicas y la tecnología, lo que significa que la empresa tenga personal, el cual tenga gran representación y análisis de datos para tener un valor agregado con información que no ha sido manejada.
2) Escala extrema de datos que supera a la tecnología actual debido a su volumen, velocidad y variedad.
3) El valor económico, haciendo que las soluciones sean asequibles y ayuden a la inversión de los negocios" **(Hopkins, 2011).**

Pero otras como ZDNet[5], una gran editorial estadounidense, lo define como las herramientas, procesos y procedimientos que permite a la organización crear, manipular y manejar grandes cantidades de datos y las instalaciones de almacenamiento **(ZDNet, 2010).**

Algunos autores eluden más a la importancia de las nuevas tecnologías y las redes sociales como punto clave de la producción de datos para elaborar una definición sobre el Big Data, sin embargo, desde mi punto de vista, aunque estos cambios en tecnología producidos por la nueva economía sea una de las principales causas de su aparición, podemos resumir su definición como la obtención, análisis y almacenamiento de grandes volúmenes de datos que no puede realizarse por medios tradicionales, y por lo tanto, se deben utilizar nuevas tecnologías para ello.

2.3 Importancia del Big Data

Según una publicación de Felipe Sevillano sobre el Big data y las Smart City, las mejores decisiones que se han tomado en la historia del mundo empresarial han estado basadas en la interpretación de unos datos disponibles **(Sevillano, 2019).** Según la revista Byte, cada día se crean más de 2,5 quintillones de bytes, por lo tanto, el 90% de los datos en el mundo han sido creados en los últimos dos años. Estos datos vienen de diversas fuentes, a través de internet la mayoría, pero también por nuestros movimientos de compras, señales GPS, el uso del teléfono móvil…etc, todos nuestros movimientos y decisiones producen datos e información sobre el consumidor y nuestros hábitos **(Navarro, 2017).**

Toda esta cantidad de datos que se está generando en la sociedad, nos está inundando de tal manera que es necesario su almacenamiento y análisis de manera

[5] Tradicional editorial estadounidense creada en 1991 por Ziff Davis.

ordenada y responsable, ya que supone hoy en día, un nuevo recurso natural a explotar y una gran oportunidad para las organizaciones que sepan utilizar y sacar buen provecho a toda esta información.

La tecnología que se encuentra detrás del Big Data es Apache Hadoop, un sistema operativo distribuido que nos proporciona la capacidad de procesar y analizar a la vez un gran volumen de datos sobre un hardware convencional. Este sistema se ocupa de una parte muy importante, como es el tratamiento de datos no estructurados, que son aquellos que no se almacenan en base de datos convencionales, pero que supone la proporción de datos mayor que produce la empresa, algunos estudios estiman que supone un 95% del total de los datos adquiridos y producidos **(Bit, 2015)**.

Sin embargo, son los datos no estructurados y su almacenamiento y análisis lo que supone un reto para las compañías. Estos datos son aquellos que no se encuentran almacenados en una base de datos tradicional **(Kyocera, 2017).** Son datos binarios que no tienen una estructura interna identificable, por lo tanto, no tienen ningún valor hasta que no son identificables y organizados. No resulta sencillo convertir los datos no estructurados en un modelo estructurado. Por ejemplo, en el caso del email, sería relativamente sencillo extraer la información y organizarla si solo tenemos en cuenta datos como la hora del envío, la persona a la que se envía o el remitente, sin embargo, el contenido del mensaje es mucho más complicado de dividir y categorizar. Alguno de los ejemplos de datos no estructurados son los archivos pdf, los audios, los videos o las publicaciones en redes sociales **(DataPrix, 2014).** Por lo tanto, para tratar estos datos, es necesario considerar las cuestiones siguientes:

- Origen
- Obtención
- Análisis de datos

A través de estos tres puntos, podremos comprender mucho mejor todo el proceso de análisis de datos Big Data. Este proceso comienza naturalmente por conocer el origen de los datos, que en muchos casos no es tan obvio como nos parece. Posteriormente, se procede a la obtención y almacenamiento de éstos. En esta fase, es muy importante contar con los soportes y materiales, que posteriormente se explican, para llevar con éxito esta fase y poder continuar con el análisis de datos. Para analizar los datos, se debe utilizar también programas especialmente diseñados para analizar un gran volumen de información, tanto estructurada como no estructurada. Cabe añadir, que existe también una parte muy importante en todo este proceso que es la interpretación de los resultados.

En esta parte final, es rol de los trabajadores especializados en Big Data es muy importante, son ellos quienes se encargan de interpretar toda la información procedente del análisis previamente realizado por las tecnologías para darle un significado y una utilidad.

3 ORIGEN, OBTENCIÓN Y ANÁLISIS DE DATOS

3.1 Origen

Los seres humanos llevamos creando y almacenando información desde el origen de nuestros tiempos, sin embargo, ha sido la gran avalancha de datos generados, capturados, almacenados y guardados por las empresas lo que ha creado el Big Data y su necesidad de introducirlo en nuestro día a día.

Empecemos imaginando los millones de usuarios que navegan por Facebook, los millones de publicaciones que se generan todos los días en Twitter o Instagram, todos los mensajes que enviamos diariamente a través de WhatsApp, los emails que envían las empresas y nosotros como usuarios al día, las llamadas telefónicas, las búsquedas en Internet, si todo esto ya nos parece una cantidad de datos difícil de recoger, almacenar y analizar, añade además muchas otras vías de producción de datos como todos los movimientos realizados con tu Smartphone, tu geolocalización, las ventas de todas y cada una de las empresas en todo el mundo, las transacciones con tarjeta, las noticias que se leen, todos los alquileres a nivel mundial, en general, todos los movimientos que realiza cada uno de los 3000 millones de internautas que hay en el mundo, el número de datos es tan grande que resulta sumamente imposible tratarlo.

Aunque la mayoría de los datos generados sean a través de la Web y las redes sociales, existen muchos otros medios de producir información que quizás no sea tan obvio para todas las personas, pero de donde también se obtiene mucha información que posteriormente se puede analizar y resultar útil para la empresa. Algunas de estas formas de origen de datos son los millones de objetos que se comunican entre sí a través de sensores, chips NFC[6], etiquetas de RFID[7], es decir, toda la comunicación de datos que se produce M2M (machine to machine) o también conocido como el Internet de las cosas.

Pero tenemos que seguir sumando muchos más datos, información que, aunque hoy en día está en la mayoría de los países digitalizada, existe mucha que sigue en papel como los millones de datos médicos que tienen los hospitales sobre cada una de las personas de cada país, datos sobre nuestros antecedentes en el registro policial, datos de

[6] Near Field Communication. Es un chip que permite el intercambio de información entre dos terminales cercanos.

[7] Identificación por radiofrecuencia. Identifica mediante un lector, sin contacto y a distancia una etiqueta o producto del almacén.

la administración pública o todos los datos producido por los sistemas de GPS, sensores en las ciudades, cámaras de video, medidores eléctricos…

Hablamos de millones y millones de bytes de información generada cada día a partir de cualquier tipo de fenómeno, desde variaciones atmosféricas hasta nuestros patrones diarios de consumo y que se difunden a través de medios muy variados, desde nuestro Smartphone como un chip introducido en una máquina.

En conclusión, según el IBM **(Barranco, 2012)**, existen más de 19 billones de dispositivos conectados a la red a escala mundial, esto significa un tráfico de datos móviles de más de 10,8 exabytes mensuales, pero si consideramos también la comunicación M2M, se cree que hay más de 30 millones de sensores interconectados y este número aumenta año por año un 30%. El crecimiento es tan grande que se estima que un 90% de los datos guardados a día de hoy han sido creados en los últimos 2 años. Ya predecía el presidente ejecutivo de Google, Eric Schmidt, "cada 48 horas generamos tantos datos como los que ha generado la civilización desde su inicio hasta 2003". Pero además de todos estos medios de producción de datos, existen muchas otras tecnologías que ayudan a organizar, almacenar y sobre todo a analizar los datos.

3.2 Análisis de datos

Posteriormente a la obtención de todos estos datos, que ya antes hemos explicado cómo son generados, damos paso a la parte de análisis. Sin duda la parte más importante junto con la interpretación de los resultados, un buen análisis de los datos puede proporcionar para la empresa una gran ventaja competitiva a la hora de tomar decisiones frente al resto.

Aunque el análisis de datos se lleva produciendo desde hace muchos años, éste ha ido evolucionando a medida que la cantidad de datos crecía. A medida que las empresas aprendan las destrezas principales para la analítica de estos datos, su ventaja competitiva será mayor y supondrá una capacidad superior para ellas. Dentro de la analítica de datos, encontramos numerosas categorías o tipos de analítica. Hoy en día estas tecnologías se centran principalmente en la integración de los datos estructurados y los datos no estructurados o semiestructurados (que son en su mayoría), es decir, en la integración tanto de los datos tradicionales y estructurados en bases de datos relacionales como la de los datos no estructurados en las bases de datos analíticas y NoSQL[8]. Esta integración de

[8] Forma y sistema de gestión de datos que difiere del modelo clásico de relaciones y análisis de datos basado en tablas. En este caso no usan SQL como lenguaje principal de consulta.

los datos facilitará y ayudará a conseguir la mayor eficacia posible, sin embargo, es muy importante que las organizaciones desarrollen una estrategia de Big Data que no sea distinta a la estrategia de datos tradicionales que ya se llevaba a cabo y conseguir una buena integración de éstos. Esto tiene una gran importancia, ya que ambas estrategias forman parte de una estrategia global mucho más grande, y aunque la cantidad de datos no estructurados es mucho más grande que la de datos estructurados, ambos deben coexistir.

3.3 Obtención y almacenamiento

Teniendo en cuenta la cantidad de datos que se generan diariamente por una persona, es normal que reflexionemos sobre la cantidad de espacio que se necesita para almacenar toda esta información. Una tendencia que va muy unida con el Big Data es el Cloud Computing, que te permite almacenar y analizar de manera sencilla y a través de la nube un gran volumen de información. Solo con recopilar valor en la nube no es suficiente para generar valor para la empresa, y analizar todos estos datos por métodos convencionales no sería posible. Es en este momento que aparece el Big Data y se base en la información recogida y almacenada en la nube para analizarlo y obtener un resultado mucho más favorable y útil **(Sabogal, 2017).**

¿El almacenamiento de todos estos datos es viable económicamente? La respuesta es clara, sí **(Torres 2010).** Si nos centramos en la empresa Amazon, podemos contratar con ellos 2 TeraBytes por 82 euros, lo cual es una capacidad de almacenamiento suficiente para muchas empresas que pueden almacenar el movimiento de su día a día, el problema viene a la hora de realizar esa lectura, leer discos duro es muy lento, pero lo que hace Google es leer 2 terabytes con 20000 discos en paralelo para en un segundo, lo mismo que se hace en computación se hace en almacenamiento.

Con estos ejemplos podemos entender mejor la rapidez en la evolución de las tecnologías de obtención y almacenamiento de datos, que se han ido adaptando al incremento sustancial de los datos que producimos, y que, por tanto, las empresas desean analizar. A continuación, se procede a explicar las herramientas principales de análisis y gestión de datos que existen actualmente en el mercado, para poder conocer de forma más detallada su funcionamiento y cómo se han adaptado a las características de Big Data.

4 HERRAMIENTAS DE ANÁLISIS Y GESTIÓN DE DATOS

Imaginamos que poseemos grandes cantidades de datos, los cuales se encuentran almacenados en grandes bases de datos y contamos con las herramientas necesarias para su análisis. En caso de partir de la hipótesis de que realizamos un análisis exploratorio, es decir, no hay expectativas de encontrar información relevante con anterioridad, sino que debemos comenzar a utilizar la minería de datos a través de campos como las matemáticas, la probabilidad o la inteligencia artificial para poder realizar una clasificación de todos los datos.

Para explicar de manera muy general el funcionamiento de la minería de datos, actualmente existe un estándar procedimental llamado CRISP-DM[9] que permite dividir el proceso de la minería en 6 fases estructuradas: Interiorizar los objetivos y requisitos, comprensión de datos, preparación de datos, modelo, evaluación y despliegue. Sin embargo, para muchos expertos, este modelo se queda obsoleto debido a las características del Big Data ya mencionadas. Por lo tanto, vamos a dividir los recursos de análisis de datos en Business Intelligence y Big Data Analytics, dos modalidades con características distintas y que analiza los datos dependiendo si éstos son estructurados, semi-estructurados o no estructurados, y dependiendo del volumen de datos.

4.1 Business Intelligence

Como lo define Jorge Conesa en su trabajo “Introducción al Business Intelligence”, es una herramienta empresarial caracterizada por transformar los datos en información que posteriormente se convierte en conocimiento. Por lo tanto, podemos definirlo como un conjunto de metodologías, aplicaciones y tecnologías dedicadas a obtener, depurar y modificar los datos que entran. De esta manera se obtiene finalmente conocimiento útil que se utiliza en la empresa para tomar decisiones, es decir, obtener una ventaja competitiva. Esta ventaja tiene como resultado la posibilidad de anticiparse en los movimientos, entendiendo mucho mejor la información del mercado y de los consumidores **(Conesa et al., 2010).**

Para entender este concepto, debemos entender que el Business Intelligence se compone de tres partes muy importantes:

[9] Metodología para proyectos de minería de datos. Es decir, un método que ha sido adoptado por organismos y empresas para llevar a cabo la minería de datos.

- **Minería de datos**: se utiliza para realizar los análisis. A continuación, se detallará más en detalle lo que es la minería de datos.
- **Data Warehouse**: su función es la de integrar todas las bases de datos que contenga la empresa.
- **Data Mart**: es una base de datos departamental y almacena los datos según el área de negocio o departamento.

4.1.1 Mineria de datos

La minería de datos es una herramienta que nos ayuda a extraer conocimiento de los datos que tenemos almacenados para posteriormente a través de la utilización de determinadas técnicas y algoritmos puedan convertirse en información útil para la empresa.

Los textos son una de las fuentes de datos más comunes y la más abundante. Podemos encontrar datos de texto en los correos electrónicos, en Whatsapp, en las redes sociales, chat en tiempo real, y lógicamente de fuentes como libros, informes, estudios y artículos de prensa entre otros muchos. La mayoría de estos textos son fuentes de datos estructurados, y desde años atrás ya se analizaban y ayudaban y su comprensión ayudaba a tomar decisiones. Existe muchos métodos y herramientas para realizar el análisis de texto, pero el más utilizado es la minería de datos.

Podemos definir como minería de datos al proceso de deducción de información de alta calidad a partir de un texto determinado **(Joyanes, 2013)**. Hoy en día, estas herramientas de análisis de texto son muy utilizadas y conocidas como parte de las suites (paquetes de software integrado) de herramientas de analítica más completas, o también se pueden obtener de forma separada e independiente, exclusivamente para el análisis de texto. Nos permite diferenciar y obtener datos no estructurados, procesarlos y crear a partir de ellos datos estructurados que puedan ser posteriormente utilizados, ya sea para la realización de informes o para procesos de análisis, pero en general nos permite obtener un valor superior que el de los datos individualmente.

4.1.2 Data Warehouse

Aunque William Harvey Inmon es considerado como el padre del Data Warehouse, fueron unos investigadores del IBM quienes le dieron este nombre. Fue creado en la década de los 90, y consiste en un conjunto de datos almacenados que pueden ser consultados por las empresas mediante tecnologías como la Minería de Datos que anteriormente se explicó. Estos datos sirven de apoyo para tomar decisiones y se encuentran ordenados, organizados por temas, son temporales y no volátiles, así es como lo definía William Harvey **(Inmon, 2007).**

4.1.3 Data Mart

Es un subconjunto de un Data Warehouse orientado al análisis, almacenamiento e integración de los datos en un departamento o área de la empresa. Posee la misma complejidad que el Data Warehouse pero se estructura de manera diferente, ya que cada departamento tiene unas necesidades distintas y una forma de gestionar y aplicar la información **(Riquelme, 2013).**

4.2 Big Data Analytics

Es una herramienta que nos permite examinar, a diferencia del Business Intelligence, grandes cantidades de datos Big Data, con el objetivo de descubrir patrones y predicciones que se puedan utilizar de forma útil en las empresas y obtener ventajas competitivas **(Camargo et. al 2015)**.

Lo que se pretende conseguir con el Big Data Analytics es poder analizar grandes volúmenes de datos y otras fuentes de datos que no se puedan analizar con Business Intelligence. Sin embargo, no debemos asociarlo solamente a grandes volúmenes de datos no estructurados, ya que también es posible realizar el análisis de bases de datos estructurados o datos relacionales. Las tecnologías que se utilizan en este caso son bases de datos NoSQL, Hadoop y MapReduce, ya que tienen la capacidad de soportar grandes volúmenes de datos. Todos estos conceptos serán explicados posteriormente ya que forman parte de las herramientas principales del Big Data **(Rodrigues, 2012).**

4.2.1 Bases de datos orientados a columnas

Las bases de datos tradicionales se encuentran orientadas por filas y son excelentes cuando queremos conocer toda la información sobre el mismo individuo o cliente, pero a medida que el volumen de datos crece y éstos empiezan a ser menos estructurados, este análisis comienza a ser más difícil y las bases de datos en columnas nos permite una mayor comprensión y tiempos de respuesta mucho más rápidos. A continuación, se explicará cómo funciona cada tipo de almacenamiento para poder ver sus diferencias.

Como explica Jorge Jara en el documento “Big Data & Web Intelligence” **(Jara, 2018),** imaginamos que tenemos la información que se representa en la primera imagen, a continuación, vamos a almacenar esta información en bases de datos orientadas a filas como se muestra en la Ilustración 2.

EmpId	Lastname	Firstname	Salary
10	Smith	Joe	40000
12	Jones	Mary	50000
11	Johnson	Cathy	44000
22	Jones	Bob	55000

Ilustración 1: Tabla de datos. Fuente: (Jara, 2018)[10]

En este caso, si se quieren analizar todos los datos de las personas que tengan un salario entre 40.000€ y 50.000€, el DBMS (Data Base Management System) tendrá que buscar en todos los datos hasta encontrar la información que de verdad necesita. En cambio, si tenemos la información almacenada en una base de datos columnares (optimizado para lograr una recuperación rápida de datos en consultas analíticas), como se ve en la tercera ilustración, en caso de querer buscar una determinada característica, el sistema podrá acceder a ello mucho más rápido sin necesidad de realizar un análisis completo.

```
001:10,Smith,Joe,40000;002:12,Jones,Mary,50000;
```

Ilustración 2: Base de datos orientados a filas. Fuente: : (Jara, 2018)[11]

```
10:001,12:002,11:003,22:004;Smith:001,Jones:002,Johnson:003,Jones:004;
```

Ilustración 3: Base de datos orientado a columnas. Fuente: : (Jara, 2018)[12]

Por lo tanto, la base de datos orientada a filas tiene que leer toda la fila para acceder al atributo que se esté buscando. Es por eso que en muchos casos se pierde mucho tiempo en consultas analíticas leyendo más datos de los necesario para satisfacer su consulta. Para ello, en una base de datos orientada a columnas, cada columna puede ser almacenada por separado, y en cualquier consulta, el sistema accederá directamente a la columna específica donde se encuentra los valores solicitados. La principal ventaja es que se reducen los tiempos y el rendimiento de las consultas lo que permite un rápido acceso a los datos. Además, cabe destacar que los elementos de análisis Big Data que se explican a continuación como NoSQL tiene como origen la orientación de las bases de datos a

[10] Analista de minería de datos en Tigo Paraguay

[11] Analista de minería de datos en Tigo Paraguay

[12] Analista de minería de datos en Tigo Paraguay

columnas, pero sobre todo en Apache Cassandra, que es la base de datos orientada a columnas más conocida y utilizada en este momento **(Garcete, 2018).**

4.2.2 NoSQL databases.

El término NoSQL fue creado en 1998 por Carlo Strozzi[13] para designar las bases de datos que no utilizaban SQL, es decir, no tienen esquemas, y tampoco almacenan datos en filas y columnas de manera uniforme **(Camargo et al., 2014).** Pero para comprender bien lo que es la base de datos NoSQL, necesitamos comprender qué significa el lenguaje SQL. Podemos definirlo como un lenguaje estructurado de consultas, como explica Platzi en su video "¿Qué es SQL y NoSQL?", es decir, es una forma de preguntar a la base de datos cierta información **(Platzi, 2015).**

Ahora bien, en la siguiente ilustración podemos ver un ejemplo de una clase de bases de datos NoSQL con las características que hemos mencionado anteriormente.

TABLA 1
EJEMPLO DE BASE DE DATOS NOSQL
CLAVE-VALOR

Clave	Valor
1	Nombre: Julio; Apellidos: Ríos; Nacionalidad: española
2	Nombre: María; Apellidos: Gutiérrez Castro; Nacionalidad: colombiana; Edad: 30
3	Nombre: Petra; Nacionalidad: italiana

Ilustración 4: Base de datos NoSQL. Fuente:(Jara, 2018)

De esta manera, al no utilizar SQL como lenguaje de consultas, no se requiere almacenar los datos en estructuras fijas o tablas.

A continuación, como se explica en el libro Big Data & Web Intelligence de Jorge Jara, se citan algunas de las principales ventajas y características de utilizar las bases de datos NoSQL para el desarrollo del Big Data **(Jara, 2018):**

- **No hay redundancia**: en las bases de datos relaciones o SQL, todo está relacionado. Vamos a ver un ejemplo para entenderlo mejor:

[13] Creador de Strozzi NoSQL Open Source Relational Database en 1998

Dentro de un restaurante, existe una gran variedad de menús distintos, y éstos a su vez tienen diferentes platos. Pero el restaurante, a la misma altura del menú, también tiene cocineros y camareros. Son parte de la relación restaurante, como se puede ver en la Ilustración 5. Al igual que el menú tiene diferentes platos, los camareros tienen diferentes atributos como: un nombre, una fecha de llegada al restaurante, la experiencia, etc.

En las bases de datos relacionales cuando definimos algo, en este caso las características de un camarero, todas las tablas tienen que ser iguales. Pero ¿qué pasa cuando quiero agregar más información sobre los camareros y no lo pensé la primera vez que cree la base? Tendríamos que cambiar, por ejemplo, el campo de edad e ir uno por uno agregando la edad. Se generaría una redundancia increíble, ya que, en el mundo real, nada es igual. Por eso se empezó a utilizar NoSQL, que nos permite tener una colección de datos de cada camarero sin que sean especialmente iguales. Es decir, no necesitamos crear relaciones nuevas cuando creamos atributos nuevos a los objetos de una colección en una base de datos no relacional **(Araujo, 2016).**

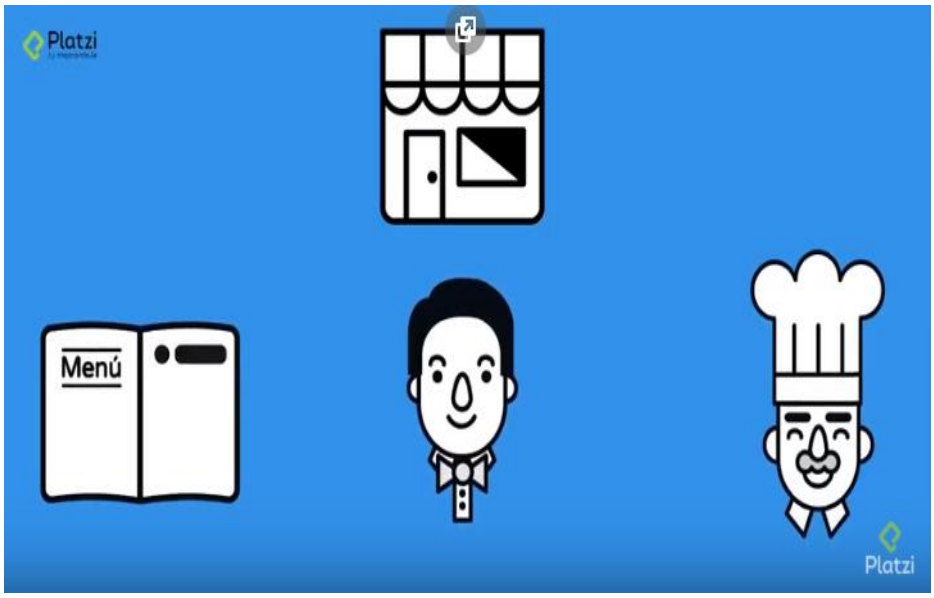


Ilustración 5: Muestra de las diferentes relaciones que hay en un restaurante para ilustrar el ejemplo de la necesidad de utilizar NoSQL. Fuente: (Araujo, 2016)

- **Es mucho más liviana**: aunque también es verdad, que algunas de las bases de datos relacionales también son extremadamente livianas. Las típicas bases de datos no relacionales son NoSQL, MongoDB, redis y CouchDB

- **Es mucho más veloz como base de datos:** al almacenar documentos en JSON[14], la velocidad de consulta es muy rápida y el rendimiento a la hora de trabajar con una gran cantidad de datos.

En las bases de datos relacionales tenemos un conjunto de tablas con información, las cuales se relacionan con un índice, en cada tabla se guarda la información de manera separada y se relaciona de alguna forma para posteriormente poder consultarlo. Todas las tablas deben estar ordenadas y contener los mismos campos e información, siempre tienen las mismas filas y las mismas columnas. Por lo tanto, en el caso de la NoSQL, al no utilizar el lenguaje SQL para la lectura de datos, tendremos que utilizar otro lenguaje, que en este caso es JavaScript. Con la base de datos no relacional o NoSQL no tenemos tablas sino colecciones donde se encuentran documentos que son objetos JSON o BSON (Binary JSON). Los documentos que encontremos se nombrarán como objetos JSON, así se organizan las bases de datos NoSQL más populares. Simplemente es estructurar los datos de una manera que cualquier programa de ordenador pueda entenderlo. Además, dentro de estos documentos pueden existir otros documentos, o simplemente más información **(Acens, 2014).**

4.2.3 MapReduce

Esta es una de las propuestas las populares para poder hacer frente al problema que nos supone el gran incremento del volumen de los datos e información. La idea de MapReduce surgió en Google en el 2004 y fue desarrollada por Dean y Ghemawat[15]. La finalidad de MapReduce es de ofrecer de manera simple, rápida, estable y resistente a fallos de poder trabajar con grandes archivos o grandes fuentes de datos, como se explica en “Una primera aproximación al descubrimiento de subgrupos bajo el paradigma MapReduce” **(Pulgar et al., 2015).**

Este paradigma se centra principalmente en dos funciones o etapas:

- **Función “Map”**:

  Toda la información llega a un clúster que segmenta el conjunto de datos de entrada en bloques que se distribuyen a los diferentes nodos de trabajo. Estos nodos de trabajo se encargarán de procesar y disminuir el

[14] El formato de archivo de JavaScript Object Notation (JSON) es un formato estándar abierto basado en texto que se utiliza para transmitir datos estructurados.

[15] Jeffrey Dean and Sanjay Ghemawat. Trabajadores en la industria Google.

problema para posteriormente pasarlo al nodo principal. Los datos generados y analizados siempre utilizarán un tipo de forma que es pares llave-valor. Durante esta fase, el mapper (encargado de la función Map) realiza alguna especie de procesamiento sobre los datos, la idea de este sistema es que se genere otro nuevo conjunto de datos, una vez más con estructura llave valor y que se envíe a la salida. Al final de esta etapa y antes de comenzar la siguiente, existe un proceso que se llama Shuffling y que consiste en ordenar los datos que has generado dependiendo de la llave que hayas asignado dentro de mapper para que tenga la misma clave y se facilite los cálculos en la función reduce.

- **Función "Reduce":**

  El trabajo de los Reducers comienza por recuperar el trabajo de los Mappers. Cada uno tendrá la tarea de procesar todos los datos asignados a una sola llave a la vez. Aquí se vuelve a realizar otro cálculo sobre los valores, pero la diferencia es que ahora tenemos la certeza de que la información que estamos analizando es valiosa y está identificada con una sola llave. De aquí se genera otro conjunto de información, que nuevamente sigue una forma de llave-valor, la cual es depositada en el archivo de salida, y podemos estar seguros de que esta información es útil **(That CS guy, 2018).**

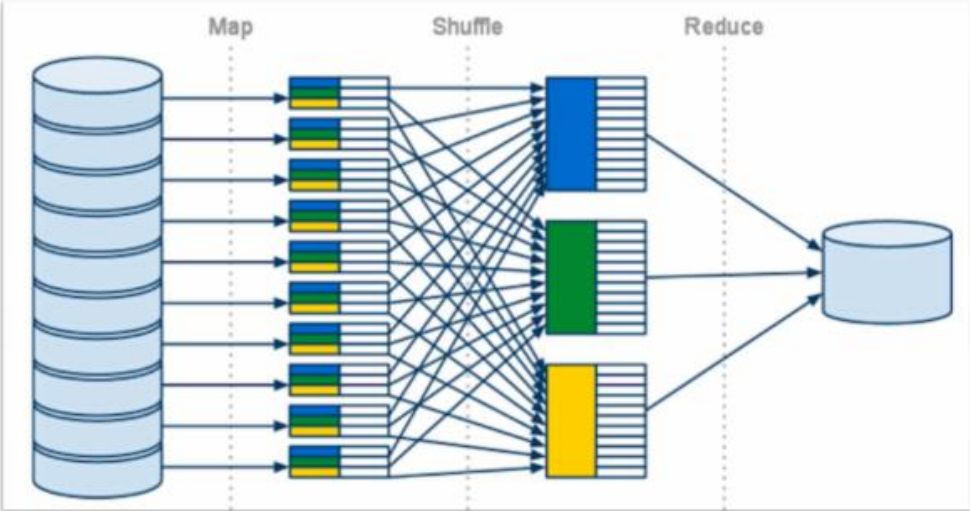


Ilustración 6: Presentación del paradigma MapReduce. Fuente: (That CS guy, 2018).

4.2.4 Apache Hadoop

Según la página oficial de Apache Hadoop, "es una biblioteca de software que permite el procesamiento distribuido de grandes conjuntos de datos a través de grupos de

ordenadores que utilizan modelos sencillos de programación. Está diseñado para pasar de los servidores individuales a miles de máquinas, cada oferta local de computación y almacenamiento" **(Apache Hadoop, 2019).** Podemos decir que es la solución tecnológica de procesamiento de datos Big Data más destacada en este momento.

El sistema Hadoop se basa en una mejora de desarrollo realizado por Google en el año 2009 del sistema MapReduce, que se centra en dos tareas, la fase Map y la Reduce, como se explicó anteriormente. Hadoop no es un programa que podamos descargar, sino que es un ecosistema de productos bajo el nombre de Apache Software Foundation. Según Mark Driver, el vicepresidente de investigación de la empresa Gartner, Apache Software Foundation es un pilar básico del ecosistema moderno Open Source, que soporta uno de los sistemas de solución de software más importantes y utilizados del mundo **(Driver, 2019)**. Los dos productos principales de Hadoop son HDFS y MapReduce, pero existen muchos otros que lo complementan y modifica como Apache Pig, Hive, Zookeeper, etc. Por lo tanto, Hadoop es un conjunto de programas que interactúan entre sí y se complementan para obtener distintas funciones dependiendo de las necesidades.

La potencia de este sistema se encuentra en que puedes realizar el análisis de todos los tipos de datos de sistemas diferentes como estructurados, no estructurados, videos, imágenes, textos, cualquier cosa que puedes imaginar. Además, se ha convertido en la mejor forma para el tratamiento de grandes volúmenes de datos en constante cambio utilizando MapReduce, como por ejemplo datos sobre el clima, sensores, M2M, o medio de comunicación basados en la web **(Cloudera, 2013).**

Para explicar cómo funciona Hadoop, es necesario comprender las características y el funcionamiento de sus tres partes principales: HadoopDistributed File Syste (HDFS), HadoopMapReduce y HadoopCommon **(Ricardo, 2012).**

HadoopDistributed File System

Es un sistema de almacenamiento de archivos tolerante a fallos, escalable y con una arquitectura distribuida. Fue creado a partir de Google FyleSystem (GFS). Los datos del clúster de Hadoop son divididos en pequeñas piezas llamadas bloques que posteriormente se distribuyen para comenzar a aplicar la función Map y Reduce en pequeños conjuntos de datos, lo que proporciona mucha mayor escalabilidad para el procesamiento de grandes volúmenes. Los elementos importantes de este clúster son **(Camargo et al., 2015)**:

- **NameNode**: solamente hay uno en el clúster. Regula el acceso a la información por parte de los clientes y lleva un control de los ficheros.

- **DataNote:** se encargan de leer y escribir las peticiones de los clientes. Los ficheros están formados por bloques que se conectan con distintos nodos cada uno.

HadoopMapReduce

Es el núcleo principal de Hadoop. Esta fase consta de dos procesos que previamente ya hemos explicado. La parte de Map, donde los datos don separados en pares de llave valor. Posteriormente se pasa a la segunda fase, Reduce, que combina los pares de llave valor en conjuntos más pequeños.

HadoopCommon

Es un conjunto de librerías que complementan la tarea de Hadoop. Aquí podemos incluir componentes muy conocidos des sistema como es Avro, un proyecto que provee servicios de serialización, o Cassandra, que es una base de datos no relacional distribuida y basada en un modelo de almacenamiento clave valor.

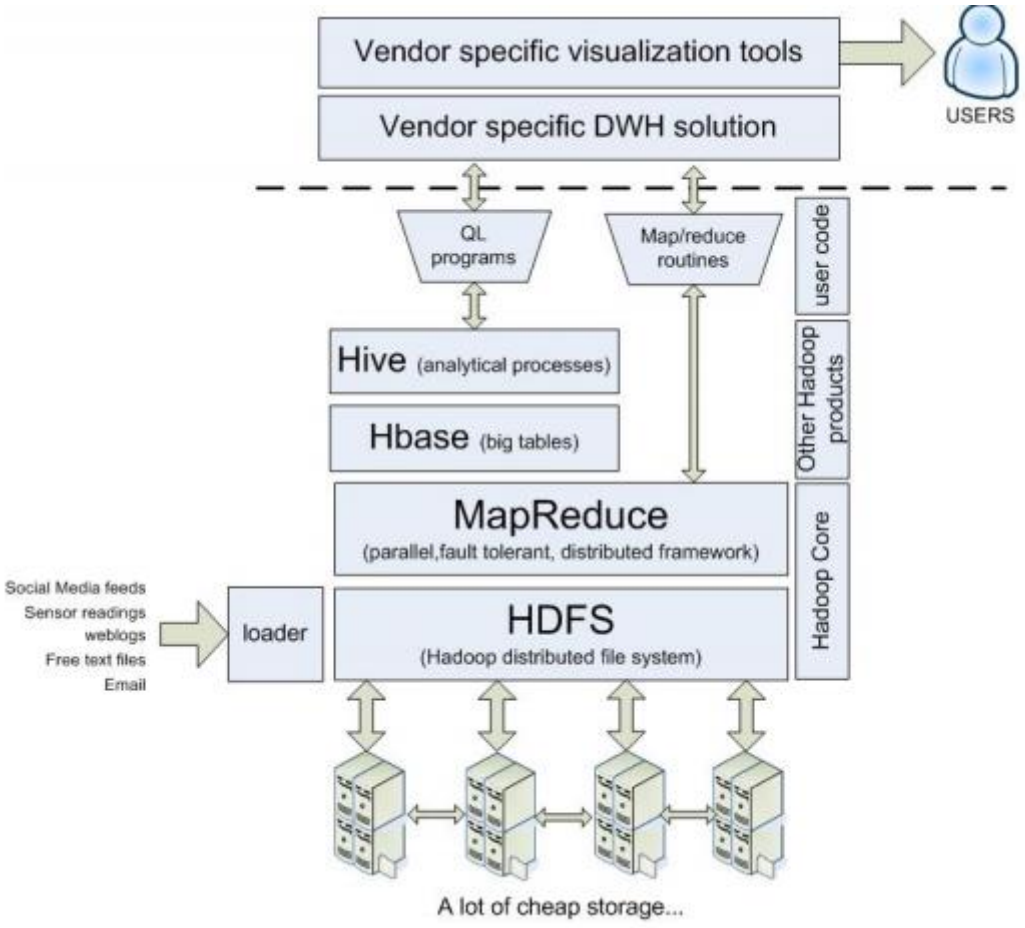


Ilustración 7: Arquitectura Hadoop. Fuente: (Jara, 2018)

El funcionamiento de Hadoop comienza cuando la información llega a un sistema de archivos llamado HDFS que recopila toda la información posible. Esta información se distribuye en nodos, que analizan estos paquetes utilizando un protocolo específico de HDFS. Se consigue fiabilidad mediante el replicado de datos a través de varios hosts, la información se almacena en 3 nodos, dos en el mismo rack y el otro en un rack diferente. A continuación, se empieza a ejecutar las funciones Map y Reduce a los pequeños subconjuntos realizados anteriormente. El resultado final son grupos de datos más pequeños y que contienen información útil. A todas estas herramientas se le puede añadir otras muchas para complementar su utilización **(Jara, 2018).**

4.3 Diferencias entre Business Intelligence y Big Data Analytics

Para poder entender mejor las diferencias entre ambos métodos, se ha procedido a la realización de una tabla resumen:

	Business Intelligence	Big Data Analytics
Velocidad	Menor velocidad de análisis.	Mayor velocidad de análisis gracias a las técnicas empleadas como las bases de datos orientadas a columnas.
Capacidad de análisis	Almacén de datos de menos capacidad.	Puede almacenar enormes cantidades de datos: Zetabytes, Petabytes…
Escalabilidad	Menor. .	Mayor.
Tipo de datos	Datos estructurados.	Datos estructurados y no estructurados.
Herramientas	Data WareHouse, Minería de datos.	Herramientas del Business Intelligence más herramientas para datos semi estructurados o sin esturar como las bases de datos NoSQL, Hadoop, MapReduce…

Tabla 1: Diferencias entre Business Intelligence y Big Data Analytics. Fuente: Elaboración propia.

Sin embargo, en mi opinión, ambos tienen muchos aspectos en común, su objetivo es ayudar a la empresa a tomar mejores decisiones y ventajas competitivas.

En el caso del Big Data Analytics, es un sistema mucho más desarrollado, sobre todo si la necesidad de la empresa consiste sobre todo en el análisis de grandes grupos de datos, será una forma mucho más eficiente y rápida, ya que es mucho más potente, es capaz de analizar un mayor número de datos y además éstos pueden estar tanto de forma estructurada, como semiestructurada como no estructurada. Mientras Business Intelligence solo podrá analizar y obtener información de las fuentes de datos estructurados, por lo tanto, tiene menor capacidad de análisis **(López, 2013).**

5 BIG DATA EN ESPAÑA

Al hablar de la implantación y utilización del Big Data en España, nos damos cuenta rápidamente en la falta de información que existe sobre las organizaciones que han implantado este sistema de obtención y análisis de datos masivos. Para comenzar, la diferencia en el porcentaje de puestos generados en España para el Big Data es bastante inferior al generado en otros países como Estados Unidos. Esta tendencia se lleva observando desde los primeros años de su aparición, en el 2014 se observaba cómo en EEUU crecía en más de medio millón los puestos de trabajo según anunciaba la empresa norteamericana Icrunchdata[16], y esta cifra ha ido multiplicándose cada año.

Sin embargo, en España la implantación del Big Data ha sido posterior y más progresiva, de hecho, hoy en día, un gran número de empresas desconocen la utilidad y la funcionalidad del Big Data. En España actualmente, el crecimiento del Big Data se sitúa en un 6% anual y, de acuerdo con el IDC, en 2021 superará los 540 millones de facturación **(Iglesias, 2018).**

Como se muestra en los estudios realizados por el INE, en 2017 solamente un 8,81% de las empresas españolas analizaban datos Big Data. Además, en su mayoría son empresas con más de 250 trabajadores o grandes empresas. Esto explica también la necesidad de analizar grandes volúmenes de datos viene ligado con el tamaño de la empresa **(INE, 2017).**

A pesar de estos datos, el 19,6% de las empresas españolas realizaron ventas mediante canales de comercio electrónico. El ecommerce cada vez está más presente en la cultura española, de hecho, casi el 20% de las compras que se realizan en España se hacen a través de Internet. Todos estos datos ayudarán y guiarán a las empresas a introducir e implantar el análisis Big Data en sus sistemas debido principalmente al aumento masivo de la información, sobretodo online.

Ha quedado claro, que entre los factores que más contribuyen al aumento de riqueza de un país se encuentra la transformación digital de las empresas, que permite finalmente poder incrementar la productividad empresarial para ser mucho más eficientes y eficaces con el servicio proporcionado. En el 2016, se realizó un estudio sobre la transformación digital en la empresa, donde también midieron la introducción del análisis Big Data en su día a día **(Ureña et al.,2017).** En este estudio se analizaron diez sectores

[16] Portal de empleo para todos aquellos que quieran trabajar en Big Data o BI en EEUU. https://www.icrunchdata.com

de la actividad empresarial y las empresas estudiadas representan un 73,2% del total del tejido empresarial español. Este estudio determinó que “un 8,8% de las pymes y grandes empresas y un 2.4% de las microempresas Big Data, siendo la geolocalización a partir de dispositivos portátiles su mayor fuente de análisis”. El sector de la información y telecomunicaciones junto con los hoteles y agencias de viajes y transporte y almacenamiento, destacan frente al resto por su mayor implicación en el uso de las tecnologías de análisis Big Data. Como ya se ha dicho anteriormente, en el sector de la información y las telecomunicaciones los datos son mucho más positivos que la media general, un 26,2% de las pymes y grandes empresas de este sector utilizan las técnicas, la tecnología y las herramientas de software para analizar todos los datos que se generan día a día tanto dentro de la empresa como fuera, lo que supone un incremento de 7,4 p.p. (puntos porcentuales) comparado con el año 2015. En el caso de las microempresas, solamente el 8,2% utiliza el análisis Big data lo que supone un incremento de 0,6 p.p. con respecto al año anterior.

En el lado contrario, se encuentra el sector de las actividades inmobiliarias y servicios auxiliares, donde solamente un 5% de las pymes y grandes empresas utilizan el Big Data en su negocio, lo que supone un descenso de 2,4 p.p con respecto al año anterior. En el caso de las microempresas este dato es aún menor y solamente un 0,8% de ellas lo utilizan, es decir, 0,8 p.p. menos que en 2015 **(Urueña et al., 2018).**

En la siguiente tabla se detalla más específicamente los datos explicados anteriormente sobre el número de empresas que utilizan software para análisis de datos Big Data por sector de actividad y por tipo de empresa.

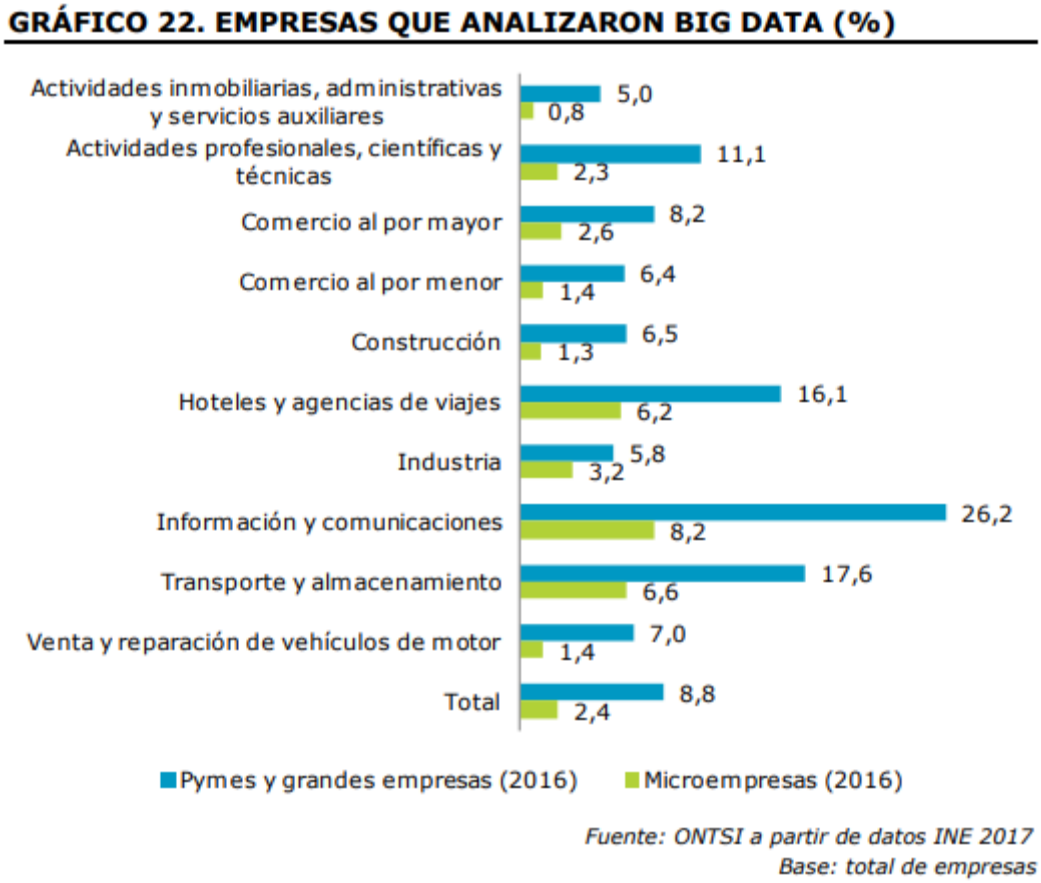


Ilustración 8: Porcentaje de empresas españolas que analizaron datos Big Data en el 2016. Fuente: ONTSI a partir de datos INE 2017

Para conocer la evolución que se ha producido en el 2016 con respecto al 2015, podemos ver en la tabla de debajo el porcentaje de empresas que han utilizado tecnología Big Data para el análisis de datos dependiendo del sector de actividad. En general, se ha producido un incremento de 0,3 p.p en el 2016 tanto en las pymes y grandes empresas como en las microempresas. Este indicador es muy positivo porque podemos ver como poco a poco se da más importancia y valor para el progreso y la eficiencia del servicio de una empresa el uso de las herramientas que se tiene al alcance para analizar los datos tanto internos de la empresa como externos, y que son de gran utilidad para ella **(Urueña et al., 2017).**

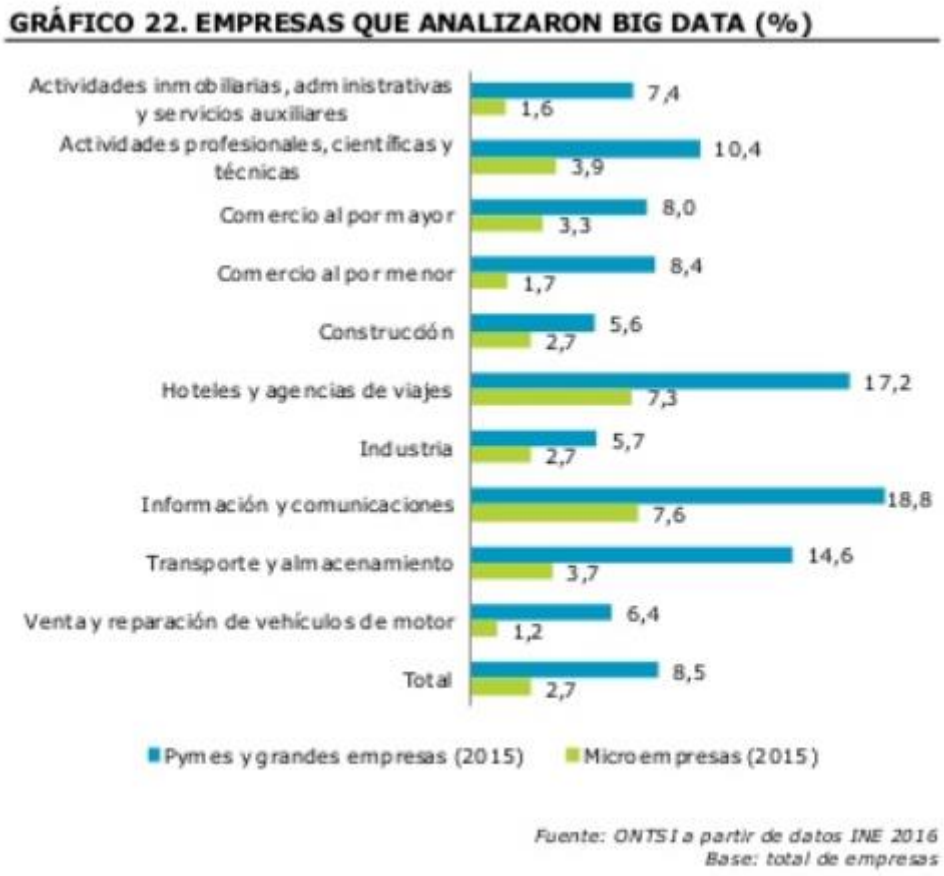


Ilustración 9: Porcentaje de empresas españolas que analizaron datos Big Data en el 2015. Fuente: ONTSI a partir de datos INE 2016

5.1 Beneficios de implantar el Big Data en la empresa

Hace algunos años, muchas empresas se dieron cuenta de que la tecnología con la que contaban no podía manejar el gran número de datos que se generaba y que almacenaban, el flujo de datos llegó a ser tan grande, que era imposible analizarlo a tiempo real. A partir de este momento, muchas empresas tuvieron que empezar a utilizar la gestión de datos Big Data, lo que aportaba a las empresas la posibilidad de unir una gran variedad de datos provenientes de fuentes diferentes en tiempo real. Esto consigue sobre todo que podamos tener una interacción más rápida y eficaz con el cliente, lo que conlleva mejores puestas en marcha de planes de marketing. Además, el análisis Big Data permite a las organizaciones conseguir un perfil más completo de sus clientes y, por tanto, experiencias mucho más personalizadas para ellos. Esta visión de 360 grados del cliente, permite también la construcción de todo un registro detallando el comportamiento de éste, y crear así un marco global y una comprensión del cliente en general **(Informática, 2016).**

Está claro de que los datos son uno de los bienes más importantes de la empresa y constituye un gran valor añadido para ésta. Tener una información y saber analizarla para

que ésta se transforme en conocimiento, supone una gran ventaja para la empresa, independientemente del sector al que pertenezcas, ya que tienes en tu posesión unos conocimientos que tus competidores no tienen y que puedes utilizar a tu favor.

Según Gartner[17], la utilización del Big Data permite a las organizaciones tener unos resultados financieros de incluso un 20% por encima de sus competidores **(Editorial, 2016).**

Hoy en día, en muchas industrias se está pensando en Big Data para solucionar muchos de los desafíos de los últimos años. De esta manera se está alcanzando una mayor eficiencia y generando un mayor valor añadido al servicio que se ofrece. Es el caso por ejemplo del sector de la salud, donde se está empezando a utilizar formas alternativas de extraer la información y mejorar el diagnóstico que se da al cliente, para poco a poco ir diseñando políticas de prevención y diagnóstico mucho más eficaces.

De la misma manera, en la industria manufacturera, se ha integrado mucha tecnología inteligente como sensores o dispositivos en sus productos para conseguir una información mucho más real y útil para mejorar y ofrecer un producto mucho más personalizado y seguro. En el caso de los automóviles, se han implantado muchas mejoras tecnológicas que miden y ofrecen información constante como el funcionamiento de todos los componentes a bordo. Todos estos datos se almacenan y ayudan a poder descubrir o identificar los problemas de fabricación o de uso mucho más rápido.

A continuación, se resumen los principales beneficios que aporta la utilización del Big Data a las empresas privadas.

5.1.1 Mejora la toma de decisiones

Disponer de un gran número de datos, sea estructurados o no estructurados, y utilizar adecuadamente las técnicas y los instrumentos de análisis Big Data para obtener de ellos información valiosa para la toma de decisiones, permitirá a la organización te tener un gran poder y una gran ventaja con respecto a los competidores. Ofrece a las empresas la posibilidad de tener una visión mucho más precisa de las fluctuaciones y de los rendimientos de cualquier recurso que lleven a cabo, permitiendo hacer adaptaciones en un momento si fuese necesario y conociendo el impacto real en su público.

[17] Empresa consultora y de investigación estadounidense de las tecnologías de la información.

5.1.2 Retroalimentación en tiempo real

El Big Data es una herramienta muy valiosa cuando la empresa debe tomar decisiones en tiempo real o a contrarreloj, ya que le permite recibir y procesar la información que entra en ese momento para conocer lo que realmente está pasando en el mismo instante en el que tomas una decisión. Se caracteriza sobre todo por ser una tecnología muy rápida y veloz que permite, por ejemplo, conocer la reacción inmediata del público al lanzamiento de un producto.

5.1.3 Conocimiento del mercado

Conocer el mercado en el que se encuentra y actúa la empresa es muy importante para poder tomar decisiones coherentes, pero también es muy importante para conocer posibles oportunidades que puede ayudarte a crecer o mejorar tu producto y servicio. Además, conocer mejor a tu público te permite también enfocar mejor las campañas de marketing y qué tipo de producto o servicio debes desarrollar.

5.1.4 Tecnología del presente y del futuro

La tecnología utilizada en el Big Data está en constante cambio y evolución, pero, sobre todo, poco a poco se va introduciendo en las empresas de todo el mundo. Supone sin ninguna duda uno de los elementos más competitivos y que proporciona mayor valor añadido a la empresa a la hora de establecer sus estrategias, por eso poco a poco es cada vez mucho más demandado por éstas, como por ejemplo los trabajadores de Business Intelligence, donde se ha vivido un gran incremento de demanda de este tipo de puesto de trabajo. Aunque su utilización es progresiva y cada vez son más las empresas que comprenden los beneficios de implantar estos sistemas, en España aún queda mucho camino por recorrer **(IEP, 2019).**

Los grandes pioneros en implantar y explotar los datos de forma masiva, en gran parte de todos los sectores son ahora claros dominantes. Empresas que supieron ver el potencial del Big Data e invertir en su aplicación, como Axciom, Google, IBM y Facebook, que siguen invirtiendo en mejoras paras descubrir nuevos usos para los datos, cómo tratarlos y transformarlo en un valor real para la empresa.

No solo es importante la ventajosa posición en que las empresas que invirtieron en Big Data en sus comienzos tienen ahora. El Instituto Global McKinsey estima que obtiene alrededor de 240000 millones de euros al año, y alrededor de 200.000 millones de euros son invertidos por la administración de la Unión Europea para mejorar cuestiones como la recaudación de impuestos, la creación de Smart Cities, y para la eficiencia energética entre otros muchos **(Gil, 2015).**

Por detrás de estas grandes empresas, se encuentran muchas compañías que están siguiendo su ejemplo y comienzan a invertir fuertemente en el desarrollo y en la innovación de estas tecnologías. En algunos casos, las empresas introducen sensores en sus productos para conocer el uso real de estos productos como por ejemplo en juguetes para niños, y de esta manera pueden ver el comportamiento del niño y desarrollar o mejorar nuevos productos mejor adaptados a ellos. Esto supone una gran ventaja competitiva para la empresa en relación con las demás. Empresas españolas como Telefónica también se han unido a este fenómeno y han comenzado a desarrollar líneas de investigación y desarrollo de sistemas Big Data. Por ejemplo, han realizado ya varios estudios sobre el turismo en ciudades como Madrid y Barcelona, en colaboración con empresas como BBVA, que consistía en analizar los datos registrados sobre terminales extranjeras a través de Telefónica y de las personas que realizaban pagos a través de servidores del banco BBVA con tarjetas extranjeras.

Sin embargo, todas estas grandes empresas coinciden en los mismo, el gran problema y la mayor dificultad es lidiar con los datos teniendo en cuenta la privacidad de sus usuarios, ya que el riesgo reputaciones al que se enfrentan en caso de realizar una mala gestión con éstos es muy grande. Y aunque en general, las empresas cada vez invierten más en Big Data y entienden mejor la importancia de integrarlo en su día a día, existen muchas oportunidades y beneficios que todavía se desconocen.

5.2 Dificultades de las empresas para implantar Big Data

Como hemos visto anteriormente, el análisis de los datos masivos o Big Data, es cada vez más importante para el éxito de una empresa. Sin embargo, como se ha explicado en el punto anterior, la mayoría de las empresas aún no han adoptado ni están interesados en adoptar esta forma de captación y análisis de datos. Cada vez existe mucha más competencia entre las empresas, y la captación y conservación de los clientes se hace mucho más complejo. De hecho, muchas empresas utilizan una gran cantidad de horas para conseguir analizar el comportamiento de sus clientes para entenderles mejor y ofrecer un servicio enfocado a su captación y retención. Sin embargo, la mayoría de ellas no cuentan con los recursos necesarios para llevar un procesamiento ágil, rápido y eficaz. ¿Cuáles son las mayores dificultades o miedos a los que se enfrenta una empresa antes de aplicar el sistema de análisis Big Data?

Los obstáculos siempre son varios, pero principalmente vamos a destacar el grado de adaptación, ya que en algunos casos son fácilmente asumibles y en otros depende de factores ajenos. El primer inconveniente es no contar con las infraestructuras adecuadas

para ello. Según el estudio realizado por la empresa Interxion, el 32% de los encuestados reconocía tener o haber tenido este obstáculo, ya que no se ha planificado de forma eficiente y tampoco contaban con demasiada información sobre cómo trabajar con Big Data de la forma más óptima posible.

El segundo obstáculo al que se enfrentan las empresas, según el 27% de ellas, se necesita una gran complejidad en la organización para llevar a cabo una correcta aplicación del Big Data. En este caso, la complejidad viene del desconocimiento de los sistemas y técnicas, pero para esto la única solución es contratar a alguien especializado en la implantación Big Data para ayudarles a entender mejor su funcionamiento **(Interxion, 2015).**

La siguiente dificultad la comparte un 26% de los profesionales, y es el desconocimiento por la normativa vigente referente al Big Data y sus límites de actuación y utilización. Existe bastante desconocimiento sobre la legislación y la seguridad de la información, por ello, es necesario también invertir en la adquisición de servicios de profesionales en la materia.

El 25% de los encuestados reconocen también que la falta de presupuestos es otro gran problema a la hora de aplicar una estrategia de análisis Big Data en su empresa. Sin embargo, estos son los mismos que argumentan la falta de información y de ayudas al desarrollo de los procesos Big Data **(Interxion, 2015).**

5.3 Principales riesgos de la utilización de Big Data

Es cierto que el Big Data proporciona enormes beneficios a través de su implantación, sin embargo, en algunos casos esto conlleva riesgos mucho más importantes que los limitados a las dificultades técnicas. A continuación, se explican los principales riesgos que nos propone la autora del libro *Big Data, privacidad y protección de datos,* Elena Gil **(Gil, 2015).**

5.3.1 Riesgo a obtener conclusiones erróneas que nadie revisa:

Una de las funciones más importantes del Big Data es poder analizar patrones de conducta para poder establecer posteriormente modelos predictivos. Pero es necesario tener en cuenta la causalidad y la casualidad, que, aunque parezcan palabras muy similares, en la práctica, puede llevarnos a errores. Existe una relación de causalidad cuando se produce una correlación entre dos variables de manera habitual. Por el contrario, cuando se produce de manera casual, decimos que esta relación es falsa y es pura casualidad. Lo que debemos entender, es que, en muchos casos, se puede producir un

fenómeno de correlación. Es por ello, que posteriormente al análisis realizado es necesario aplicarle una parte subjetiva y estudiar realmente si existe relación o si por el contrario, ha sido una mera casualidad y las variables no tienen conexión. De hecho, cuantos más datos se analicen, más probabilidad existe de encontrar correlaciones entre ellos o entre varias variables. Esto significa, que se puede estimar de manera errónea, si no se realiza un estudio humano posterior, el resultado del análisis, y por tanto, la información proporcionada no es útil.

5.3.2 Riesgo de la toma de decisiones de forma automatizada.

La mayoría de las operaciones que se realizan en Internet se hacen de forma automatizada, donde el ser humano no ha intervenido para nada, salva para establecer los parámetros necesarios para la toma de la decisión. Tomar las decisiones de forma automatizada está llevando a muchas empresas a llevar a cabo proyectos sin saber exactamente por qué se han tomado. En el marketing, un error de estos puede suponer una mala publicidad, por ejemplo, sin embargo, llevar a cabo una decisión equivocada en el sector bancario, puede hacer perder millones de euros.

5.3.3 Riesgo para la privacidad de las personas.

Es uno de los mayores riesgos y retos a los que las empresas deben enfrentarse. A continuación, se va a hablar más detalladamente sobre la legislación actual de protección de datos y el impacto que tiene en las empresas.

5.4 La legalización del Big Data

Hoy en día tenemos acceso a un gran número de tecnologías que están revolucionando el mundo y la forma en que éste funciona. Hemos llegado a un punto en el que los datos se han convertido en uno de los bienes más preciados. En un momento de la historia que se crea más datos que nunca antes, poder recogerlo, almacenarlo y tratarlo se ha convertido en un desafío. Sin embargo, muchas empresas ofrecen servicios gratuitos a cambio de poder tener acceso a esta información y utilizarlo para numerosos fines y de muchas maneras distintas. En muchos casos, esta información es simplemente conocer el comportamiento o los patrones de consumo de los clientes, que ya proporciona un gran valor a la empresa, que podrá diseñar y crear una publicidad dirigida mucho más efectiva que la de los competidores.

Todas estas nuevas tendencias vienen acompañadas de muchos riesgos. Uno de los principales es la realización del análisis masivo de datos y cómo interviene e influye a la privacidad de las personas. De hecho, el problema viene cuando los avances y la

tecnología crecen de manera tan rápida que en muchos casos a la ley no le da tiempo a cambiar y adaptarse, por lo que no son capaces de dar una respuesta a estos nuevos problemas que se plantean hoy en día. A continuación, centraré mi análisis en los principales riesgos que conlleva el Big Data, y más en profundidad sobre la privacidad y la protección de datos **(Gil, 2015).**

Para explicar los límites y problemas legales sobre el Big Data, y antes de analizar el nuevo Reglamente de Protección de Datos (GDPR) que se implantó en el 2018, es necesario hablar y conocer los siguientes conceptos.

5.4.1 ¿Qué son los datos de carácter personal?

Según el BOE, en diciembre del 2018, se estableció en su artículo 3 la siguiente definición: "cualquier información concerniente a personas físicas identificadas o identificables" **(BOE, 2018)**

De esta manera, una persona es identificable cuando podemos determinar su identidad en base a cualquier información especificada como la fisiológica o la psíquica. Los datos de carácter personal no se refieren únicamente a nuestro nombre y apellidos, sino que cada vez se añade más información como nuestra dirección, el número de teléfono o los "me gusta" que damos en Facebook. Cualquier forma o información que nos identifique forma parte de nuestros datos de carácter personal.

En la definición proporcionada anteriormente, debemos analizar detenidamente cada parte para entender más en detalle qué protege la ley y cómo lo protege.

"Cualquier información"

Esto quiere decir, que todas las normas comunitarias de protección de datos abogan por una protección completa, en el sentido amplio de información. Además, lo que se entiende por información de carácter personal va evolucionado y ampliándose con el paso del tiempo y la evolución de las tecnologías. Por último, destacar que esta información no tiene por qué ser verídica ni que se contemple únicamente en bases de datos, de hecho, ésta puede ser incorrecta y puede contenerse en cualquier formato como un texto o imágenes.

"Persona identificada o identificable"

Una persona es identificada cuando se conoce a quién pertenece una información, y es identificable cuando, aunque aún no se conozca a quién pertenece esa información, es posible hacerlo.

5.4.2 Reglamento general de protección de datos (RGPD)

El próximo Reglamento General de Protección de Datos (RGPD) entrará en vigor en mayo del 2019, y lo que busca es mejorar el trato de datos por terceros, más concretamente, busca que cada persona tenga que autorizar de forma explícita el propósito de ceder sus datos. Tanto para las compañías que manejan datos personales a gran escala como para las pequeñas empresas, tendrán que tener una gran transparencia en la gestión de la información. A diferencia de la anterior ley vigente, a partir de ese momento, el usuario tendrá pleno control de sus datos y si decide el procesamiento de éstos, teniendo la capacidad de disponer y elegir la gestión que se le da a los mismos **(Prometeus, 2019).**

Debemos saber que la protección de las personas físicas en cuanto al derecho de la protección de datos de carácter personal, se establece en el artículo 8.1 de la Carta de los Derechos Fundamentales de la Unión Europea y en el art 16.1 del Tratado de Funcionamiento de la Unión Europea. En España, la protección de datos de carácter personal se encuentra en el art 18 de la Constitución Española de 1978, que garantiza “el derecho al honor, a la intimidad personal y familiar y a la propia imagen”. Aun así, en mayo del 2016 se publicó un reglamento específico en la Unión Europea, que se renovó en el 2018 sobre la protección de datos en el seno de la Unión Europea. Este reglamento se aplica de manera general a todos los estados miembros y es obligatorio en todos sus elementos. Cada día, debido a la evolución de las tecnologías y del crecimiento de datos personales, se requiere un marco más sólido y coherente para la protección de datos en la UE, es por ello que se creó la RGPD. La legislación española también ha tenido que adaptarse a ello a través de un Real Decreto-ley y que tiene como objetivo complementar y aclarar los conceptos del RGPD para ofrecer sobre todo muchas más garantías al consumidor en derechos digitales **(Serrano et al., 2018).**

5.4.3 La normativa de protección de datos y el Big Data

Como ya hemos visto antes, esta normativa es aplicable cuando las personas físicas son identificadas o identificables, por lo tanto, cuando los datos no identifican a una persona en concreto, no se aplica esta regulación. La tecnología Big Data, utiliza técnicas de anonimización para que los sujetos no sean conocidos ni identificables, y en consecuencia, estos datos ya no se consideran datos de carácter personal y no se aplicaría la RGPD.

El Big Data desafía un poco las normas de protección de datos, aplicando técnicas de re-identificación que permiten conocer la identidad de los sujetos a partir de unos datos que habíamos considerado anónimos en el pasado. Por lo tanto, estas técnicas de

anonimizacion a veces no son suficientes con la llegada del Big Data. Por lo tanto, el Big Data amenaza y desafía a esta normativa en muchos sentidos. A parte de que no se encuentra adaptada a los medios técnicos y desarrollos de hoy en día, la normativa confía mucho en el consentimiento informado del individuo para poder tratar sus datos, sin embargo, nos damos cuenta de que esto es bastante subjetivo ya que la mayoría de los individuos tan siquiera leen las políticas de privacidad antes de aceptarlas. Si bien la anonimizacion se presenta como la mejor solución para tratar los datos de los consumidores sin conocer su identidad, se han conocido muchos casos de bases de datos que se han vuelto a re-identificar a los sujetos, lo que supone un gran riesgo para la privacidad de todos ellos.

5.4.4 ¿Qué deben hacer las empresas?

Cualquier empresa que ofrezca un servicio en la UE, está obligada a cumplir con el Reglamente de Protección de Datos. La transferencia de datos entre los países de la UE está permitida, en caso de países fuera de esta zona, se debe tomar restricciones y medidas añadiendo clausulas según los modelos de contrato de la UE. De hecho, también se aplica a esta ley las empresas que estén fuera de la UE pero que recojan datos de ciudadanos europeos, la ubicación geográfica de la empresa no importa.

Es importante que las empresas realicen una clara identificación de los procesos donde existe el tratamiento de datos personales. Además, el GDPR requiere del consentimiento expreso de la persona para registrarse y dejar sus datos. Es importante que la empresa desarrolle técnicas especiales para que dicho consentimiento quede expresado claramente por el consumidor o cliente. En el caso de las empresas y organismos del sector público, deben obligatoriamente contar con una persona especializada en la protección de los datos, para poder reconocer la cantidad y a calidad de estos, tanto si son los de sus propios empleados como los de sus clientes. A este perfil se le llama DPO (Delegado de protección de datos) **(Interxion, 2017).**

A continuación, se detallan algunos de los puntos más importantes que debe cumplir la empresa con la GDPD:

Consentimiento

Se debe tener el consentimiento de la persona para poder almacenar y utilizar sus datos, explicando además para qué serán utilizados. Este consentimiento se debe hacer de forma “expresa, precisa e inequívoca”. Además, corresponde al titular de los datos determinar cuáles quiere que sean tratados y registrados y qué uso se les puede dar.

Derecho al olvido

Los ciudadanos europeos pueden solicitar el olvido y borrado e sus datos personales si estos lo desean al responsable del tratamiento de los datos, y además, podrá solicitar la prohibición de compartirlos con terceros. En este caso, hay que tener en cuenta, de que hoy en día, aunque un prestador de servicios de la sociedad de la información retire tu información de su base de datos, estos posiblemente sigan en Internet durante mucho tiempo ya que esta información se almacena en los buscadores y en la memoria denominada caché, y permite que ésta aparezca en la web sin control alguno.

El derecho al acceso

Si un ciudadano europeo lo solicita, podrán pedir a la empresa el acceso a los datos que poseen, al igual que ser informados sobre su finalidad y dónde se encuentran almacenados. Esto está relacionado con un elemento que se buscaba incrementar con la creación e estas nuevas normas que es la transparencia de la información. De hecho, es en el artículo 15 y ss. Del RGPD donde se regula los derechos al acceso, rectificación, derecho al olvido, limitación y oposición **(Blanco, 2018).** El interesado puede ejercer estos derechos frente al poseedor de un fichero de datos con el fin de conocer sus datos personales y el uso que se les está dando.

Notificación obligatoria en caso de filtración de datos

Las empresas cuentan con 72 horas para notificar a la autoridad pertinente en caso de que se produzca un problema en su sistema y se ponga en riesgo la seguridad y los derechos de las personas.

El riesgo de no cumplir con el reglamento

Existen diversos rangos de multas si no se cumple con la ley. En los casos más extremos, se puede llegar a obtener sanciones por unos 20 millones de euros y el 4% de la facturación del año anterior en el caso de las empresas, según lo que sea más alto. Para infracciones menores, las multas suelen ser la mitad de los establecido anteriormente, alrededor de 10 millones de euros o un 2% de la facturación del año anterior **(Interxion II, 2017).**

6 ESTUDIO DE MERCADO

6.1 Objetivo del estudio

Como ya se ha hablado anteriormente, el objetivo de este estudio era conocer la opinión tanto de los usuarios que consumimos de forma online en nuestro día a día como de la importancia que le da la empresa al tratamiento de datos Big Data.

En este primer estudio, a través de un cuestionario lanzado a las personas residentes en Valladolid que utilizan tecnologías como un ordenador o un teléfono móvil, se pretende conocer el grado de importancia que tiene para ellos la utilización de sus datos e información generada por la utilización y la navegación online por las empresas y la forma en que éstas los gestionan y los tratan.

La razón de esta hipótesis se encuentra en que hoy en día, los usuarios no tenemos suficiente conocimiento sobre la utilización de nuestros datos y rastreos a través de la web, por un lado, no conocemos realmente la cantidad de información que generamos con el simple hecho de hacer una búsqueda online o entrar en una página, y por otra parte, tampoco conocemos cómo las empresas gestionan esta información y si realmente supone una gran utilidad para la empresa.

Para complementar el estudio, se ha querido conocer también la opinión de algunos expertos en el tema, que trabajan en empresas donde la utilización de datos Big Data supone su día a día, conocer qué grado de importancia tiene para ellos la utilización de estas técnicas de análisis y almacenamiento y cómo esto ha ayudado a la empresa.

6.2 Descripción de la zona estudiada

Para determinar la población de estudio y la muestra, se ha de determinar primero el entorno y el área donde se va a realizar el análisis. He querido centrarlo en Valladolid, eso quiere decir que la población es de 519.851 en 2018 según el INE. **(INE, 2018).**

La encuesta se ha realizado a las personas residentes en toda la provincia de Valladolid, lo que incluye todas las zonas periféricas y todos los pueblos que se encuentren en esa área geográfica.

6.3 Análisis del estudio

Ahora que ya conocemos más detalladamente la zona y la población de estudio, el siguiente paso es delimitar la población de estudio para encontrar la muestra representativa para nuestro análisis.

Muestra de los consumidores

Para determinar la población de estudio de los consumidores, se ha obtenido del Instituto Nacional de Estadística la población comprendida entre los 18 y los 65 años de edad en el total de la provincia de Valladolid. He decidido delimitar la población en este rango de edad ya que se ha estimado que son las personas comprendidas entre esta edad los más propensos y con más posibilidad de consumir a través de Internet o utilizar Internet en su día a día. A este número de individuos se les ha vuelto a delimitar aplicando un porcentaje obtenido también del Instituto Nacional de Estadística, sobre las personas que han comprado en los últimos 3 meses en Internet.

Por lo tanto, para definir la población de estudio (N), se han efectuado los siguientes cálculos:

- Población de la provincia de Valladolid comprendida entre 18 y 65 años: 326.956
- Porcentaje de personas que han comprado en Internet en los últimos 3 meses en Castilla y León: 41,8%. Aplicamos este porcentaje ya que no se ha podido encontrar un porcentaje aplicado solamente a la ciudad de Valladolid.
- Aplicando este porcentaje, la población de estudio resultante (N) sería: 136.667

La encuesta siguiente se ha realizado en el mes de diciembre de 2018 a nivel provincial, en Valladolid. La en cuestación se ha llevado a cabo a través de la página web Google Encuestas [18] donde los encuestados tenían que responder a 10 preguntas que se les hacía acerca de su conocimiento y opinión sobre el Big Data, para poder conocer más en detalle cómo influye en ellos a la hora de consumir.

Para la determinación del tamaño de la muestra, emplearemos un muestreo aleatorio simple y no estratificado, puesto que no disponemos de medidas de variabilidad en materia de consumo a través del comercio electrónico dentro de la propia población de estudio.

Ahora que ya conocemos el tamaño de la población de estudio, procedemos a calcular el tamaño de la muestra (n) para realizar la encuesta. Para ello, primero debemos identificar si se trata de una muestra finita o infinita. Según las reglas de estadística, una población es finita cuando está formada por un conjunto menor a los 100.000 elementos, está formado por un conjunto de personas que tienen unos atributos y características comunes y que tiene una cantidad limitada de integrantes, es decir, que es sencillo

[18] Se puede acceder desde esta URL https://forms.gle/xBQJb2hFMLEH6dJL9

identificarlos. Por el contrario, es una población infinita cuando la muestra es mayor a 100.000.

A continuación, se procede a aplicar la fórmula de la población infinita para calcular el número de encuestas necesarias para que el estudio sea representativo.

$$n = \frac{k^2 \times p \times q}{e^2}$$

Donde:

K: Constante que depende del nivel de confianza que se asigne. Para esta investigación se ha utilizado un nivel de confianza del 95%, por lo tanto, k será igual a 2.

e: Es el error muestral deseado. Para este estudio hemos admitido un error de 0,07.

P: Probabilidad de que los individuos consuman comercio electrónico al menos una vez durante los últimos 3 meses: 41,8%

Q: Probabilidad de que los individuos no consuman comercio electrónico durante los últimos 3 meses.

De acuerdo con el informe realizado por el INE sobre la utilización de productos TIC por los residentes en Castilla y León (2017), obtenemos los datos relativos a estas variables que son p=41,8% y q=58,2%.

n= Tamaño de la muestra

Sustituyendo esos datos en la fórmula anterior, quedaría de la siguiente forma:

$$n = \frac{2^2 \times 0{,}418 \times 0{,}582}{0{,}07^2} = 198.59 \approx 199$$

Eso quiere decir que, teniendo en cuenta el posible error del 7% utilizado en e, se han de realizar un mínimo de 199 encuestas entre los consumidores para que el estudio sea representativo entre los consumidores.

6.4 Trabajo de campo

Este estudio de mercado se realizó en el mes de diciembre de 2018 a nivel de la ciudad de Valladolid. La encuesta se ha realizado a través de Google Drive, donde todos los encuestados podían responder a las preguntas que se proponía. El enlace a la encuesta se denomina "Encuesta para los consumidores. Utilización del Big Data"[19]. A través de este formato, he podido distribuir la encuesta por redes sociales y por la universidad, para poder dirigirme a mi público objetivo. Esta encuesta se podía responder en 3 minutos tanto con el ordenador como por un Smartphone y se compone de preguntas creadas de forma que progresivamente se va conociendo el nivel de conocimientos sobre el Big Data de los usuarios y si le dan importancia a la hora de consumir.

A continuación, se presenta el modelo de cuestionario de la encuesta propuesta y realizada por los consumidores. En el anexo se incluye imágenes sobre cómo los consumidores visualizaban la encuesta a través de la plataforma online.

1. **¿Qué te parece que las empresas utilicen información sobre los hábitos de los usuarios para incrementar sus ventas?**
   - ☐ Me parece muy bien porque también nos beneficia como consumidor, ya que nos ofrecen productos y servicios mejor adaptados a nuestras necesidades.
   - ☐ No me parece muy bien porque muchas veces obtienen estos datos sin nuestro consentimiento.
   - ☐ Me parece fatal, aparte de no saber qué están haciendo con nuestra información, la obtienen de manera gratuita y sin consentimiento.
   - ☐ No tengo opinión en ello.

2. **¿Cambiarías o dejarías de consumir en una empresa si conocieras que ésta no utiliza tus datos de forma legal y sin tu consentimiento?**

   - ☐ Sin ninguna duda dejaría de consumir en ella. No me parece ético y no pondría mi granito de arena para hacerlo.
   - ☐ Me lo pensaría si tengo más opciones donde consumir.
   - ☐ No le doy importancia, mientras me ofrezca un buen servicio o producto.

[19] Se puede acceder desde esta URL: https://forms.gle/xBQJb2hFMLEH6dJL9

3. **¿Te preocupa que las empresas tengan tantos datos sobre ti, te conozcan tan bien y no sepas qué están haciendo con esos datos?**
   - ☐ Me preocupa mucho.
   - ☐ Me preocupa algo.
   - ☐ Me preocupa poco.
   - ☐ No me preocupa para nada.

4. **Señala cuáles de las siguientes formas de recogida de información conoces actualmente.**
   - ☐ Soportes asociados a una página de compra.
   - ☐ Las cookies.
   - ☐ Tarjetas bancarias, tarjetas de fidelización.
   - ☐ Smartphone.
   - ☐ Redes sociales.
   - ☐ Conexión wifi.
   - ☐ Smart cities o ciudades inteligentes.

5. **Al utilizar cualquiera de las opciones anteriores, ¿Piensas en la cantidad de información que estás dando de manera gratuita a las empresas?**
   - ☐ Sí, constantemente. Intento impedirlo teniendo desactivado los ajustes debidos y no dando mi consentimiento para ello.
   - ☐ Lo pienso alguna vez y rechazo que obtengan esta información cuando puedo.
   - ☐ No se me pasa mucho por la cabeza.
   - ☐ Nunca lo he pensado.

6. **En general, ¿qué tal de familiarizado estás con el concepto Big Data?**
   - ☐ Muy familiarizado, comprendo su uso y cómo las empresas lo utilizan.
   - ☐ Bastante familiarizado, he leído algo sobre ello, pero no conozco muy bien la idea.
   - ☐ Poco familiarizado, he escuchado el concepto pero no lo comprendo mucho.
   - ☐ Nada familiarizado, no conozco nada sobre el tema.

7. **¿Consideras que el Big Data (recogida y análisis de datos) proporciona un gran valor para la empresa?**
   - ☐ Sin ninguna duda, aporta una información muy importante y les da una gran ventaja competitiva.
   - ☐ Bastante, pero no es de gran valor, no supone una gran ventaja frente al resto.
   - ☐ Un poco sí, pero muchas veces es una pérdida de tiempo y dinero.
   - ☐ Para nada.

8. **¿Crees que es necesario para las empresas?**
   - ☐ Totalmente, hoy en día aún más, donde es indispensable para una empresa analizar un gran volumen de datos y conocer al consumidor.

- Bastante importante pero no necesario. Las empresas pueden sobrevivir sin utilizar el Big Data.
- Depende de la empresa, pero no lo encuentro necesario.
- No es necesario para ninguna empresa.

9. **Edad:**
   - Entre 18 y 24 años.
   - Entre 25 y 30 años.
   - Entre 31 y 60 años.
   - Más de 60 años.

10. **Sexo:**
    - Femenino.
    - Masculino.

6.5 Análisis de los resultados obtenidos

El procedimiento llevado a cabo en el análisis de la encuesta a los consumidores fue el siguiente: los datos se exportaron a un fichero CSV, que contenía un documento EXCEL donde se encontraba recogida toda la información concerniente a la encuesta. El análisis se completó con las gráficas que se crean directamente con los Formularios Google.

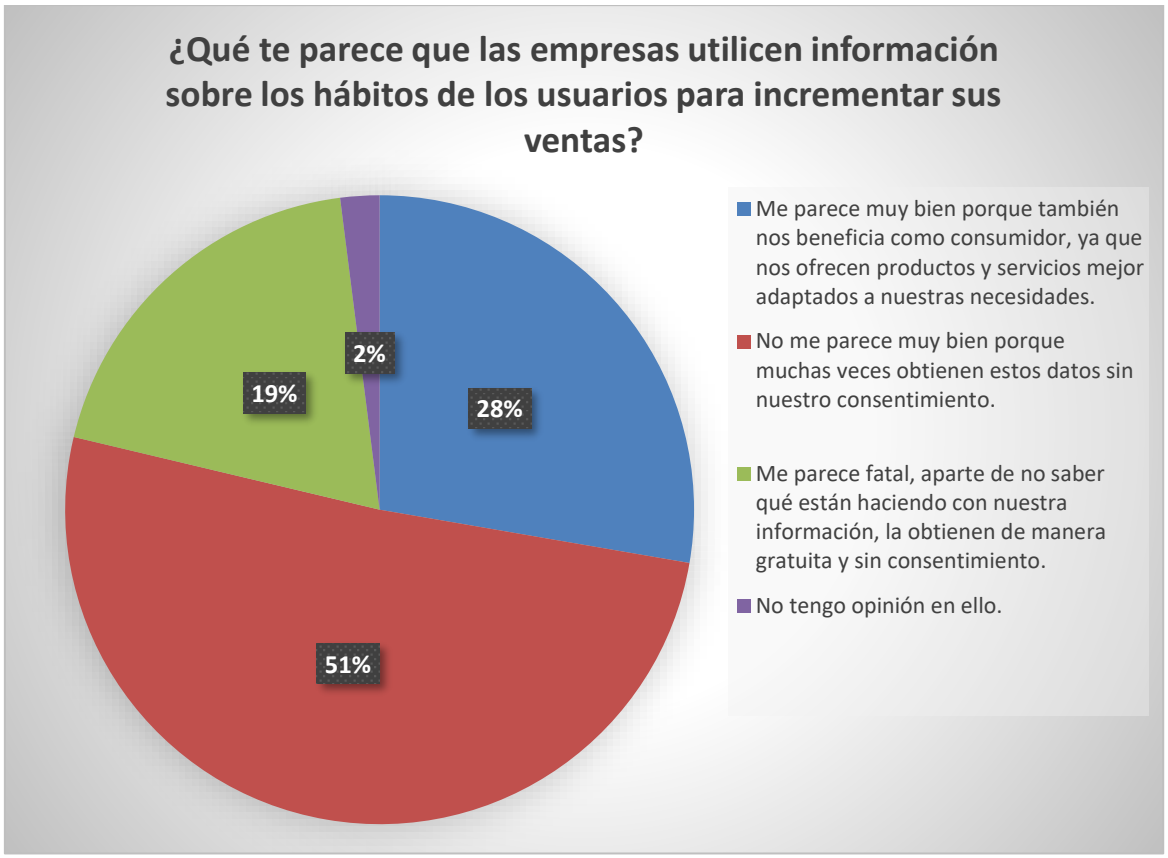
6.6 Interpretación de los resultados

Ahora que ya conocemos cuál ha sido el proceso que se ha llevado a cabo para el desarrollo de las encuestas, a continuación, se procede al análisis de los resultados donde se han añadido las conclusiones que puedo obtener de ello.

Los gráficos que se adjuntan son los creados por los Formularios de Google. A partir de ellos se procede a analizar los resultados añadiendo mis conclusiones.

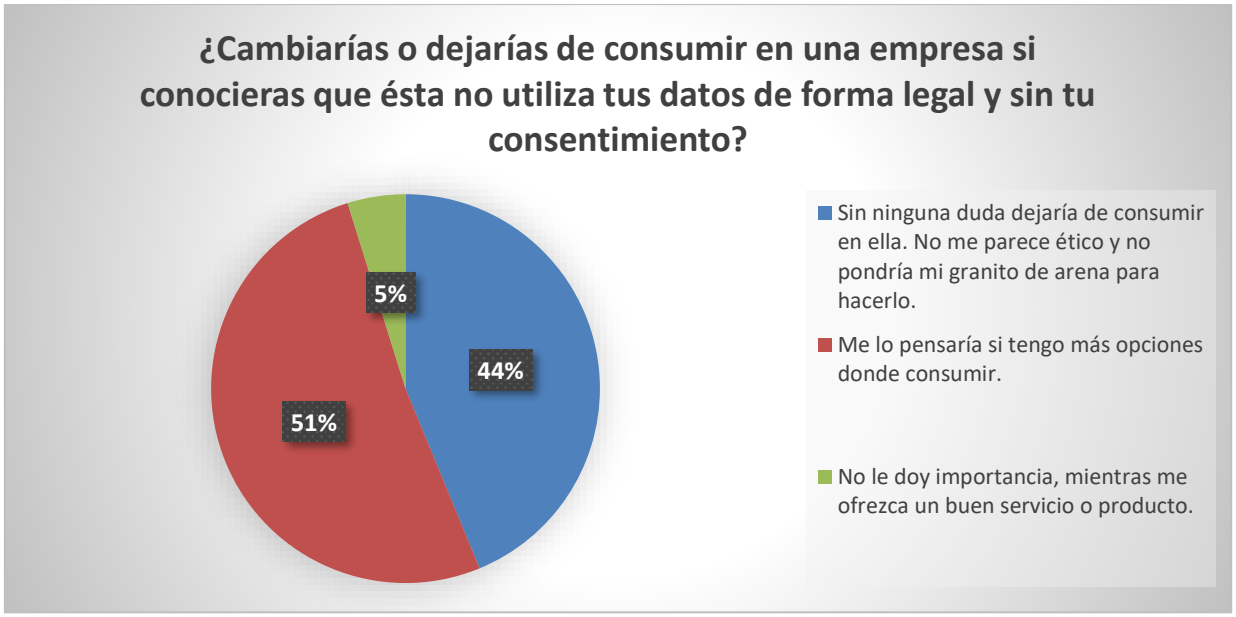
La primera pregunta que se propone en el formulario pretende conocer la opinión de los consumidores respecto al conocimiento de su información por parte de las empresas con fines comerciales, sin entrar en términos de Big Data.

A alrededor del 70% de los consumidores les parece mal que las empresas utilicen su información para incrementar sus ventas. Solamente 69 de los encuestados (27,7%) está de acuerdo con ello porque piensa que ellos también están beneficiados al recibir productos mejor adaptados y personalizados a sus necesidades, o simplemente muchas veces, más justos.



Gráfica 1:*Opinión de las personas sobre la utilización de su información por las empresas***. Fuente:** Formularios Google

En el siguiente gráfico podemos ver cómo un 95% de los consumidores dejarían de consumir en una empresa que obtiene información sobre sus consumidores de forma ilegal. Solamente un 5% seguiría consumiendo a pesar de ello.

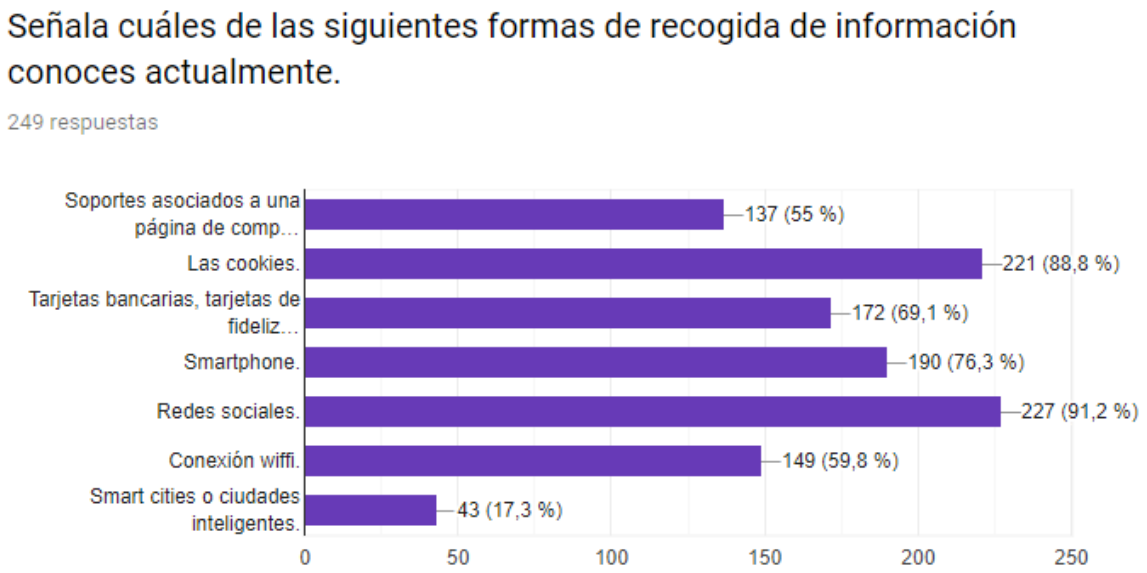


Gráfica 2: *Opinión sobre el comportamiento a la hora de consumir si una empresa utiliza tus datos de forma ilegal o sin tu consentimiento***. Fuente:** Formularios Google

Sin embargo, solamente el 34% está realmente preocupado por la posesión de su información por parte de las empresas. Más de un 50% lo tiene presente pero no le da gran importancia. Solamente un 16% de los encuestados está muy poco o nada preocupado por ello.

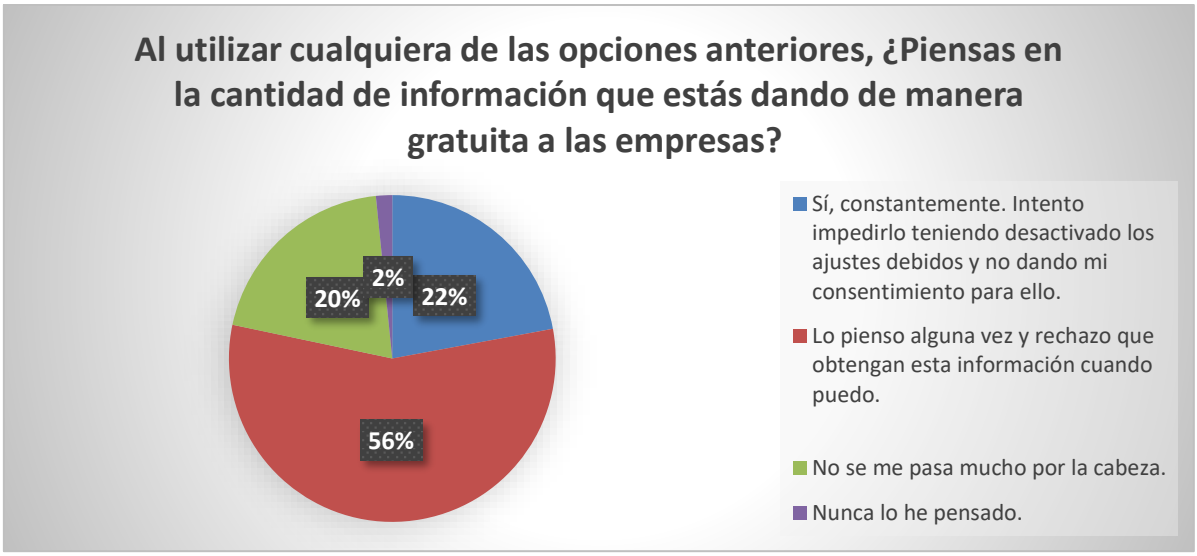
Gráfica 3: *Preocupación de los usuarios sobre la posesión de su información por parte de las empresas*.
Fuente: Formularios Google

La siguiente pregunta propuesta a los usuarios es de opción múltiple. En ella, tenían que contestar qué formas de las propuestas conocen como fuente de recogida de información. Podemos ver como la mayoría, con casi un 92% de los encuestados, saben que las redes sociales son una fuente para la recogida de información. Seguido por las cookies y nuestro Smartphone. Un poco más del 50% reconocen saber que a través de su conexión wifi y los soportes asociados a su proceso de compra. Solamente un 17,3% conoce que las SmartCities son una forma de recogida de información.



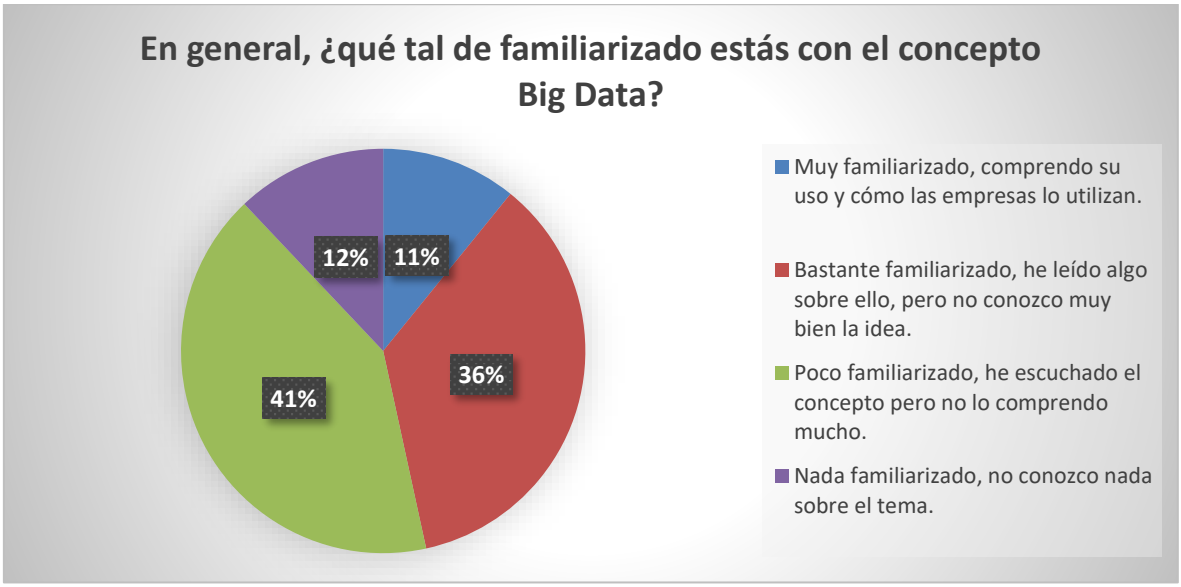
Gráfica 4: *Conocimiento de los consumidores sobre las diferentes fuentes de recogida de información.*
Fuente: Formularios Google

Un 20% de las personas que utilizan las formas propuestas anteriormente no piensan en la información y en la huella digital que están dejando y dando de forma gratuita. Un 22% lo piensa constantemente e intenta realizar todos los ajustes necesarios para registrar la mínima información. Un 56% lo piensa de vez en cuando pero no le da mayor importancia.



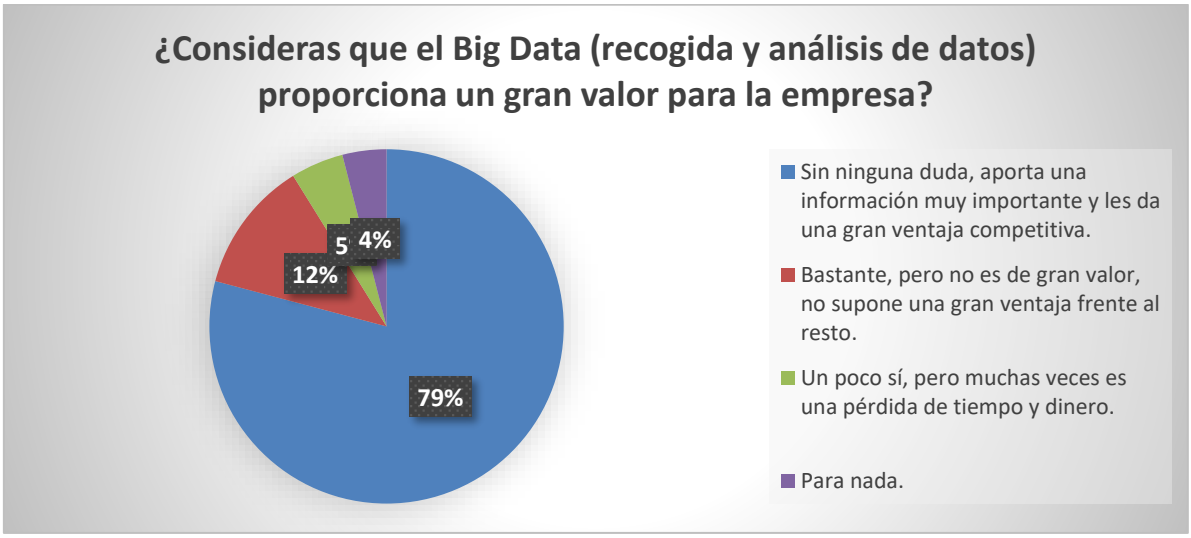
Gráfica 5: *Frecuencia con la que los consumidores piensan en la información que están dejando al utilizar algún sistema electrónico*. **Fuente:** Formularios Google

De hecho, podemos ver que solamente un 10,8% de los encuestados están muy familiarizados con el concepto Big Data. Un 35,7% ha escuchado hablar de ello, pero no tienen mucha información ni comprenden totalmente qué es lo que es y cómo funciona. Un 41,4% de los encuestados está muy poco familiarizado y solamente un 12% no ha escuchado nunca hablar sobre ello.



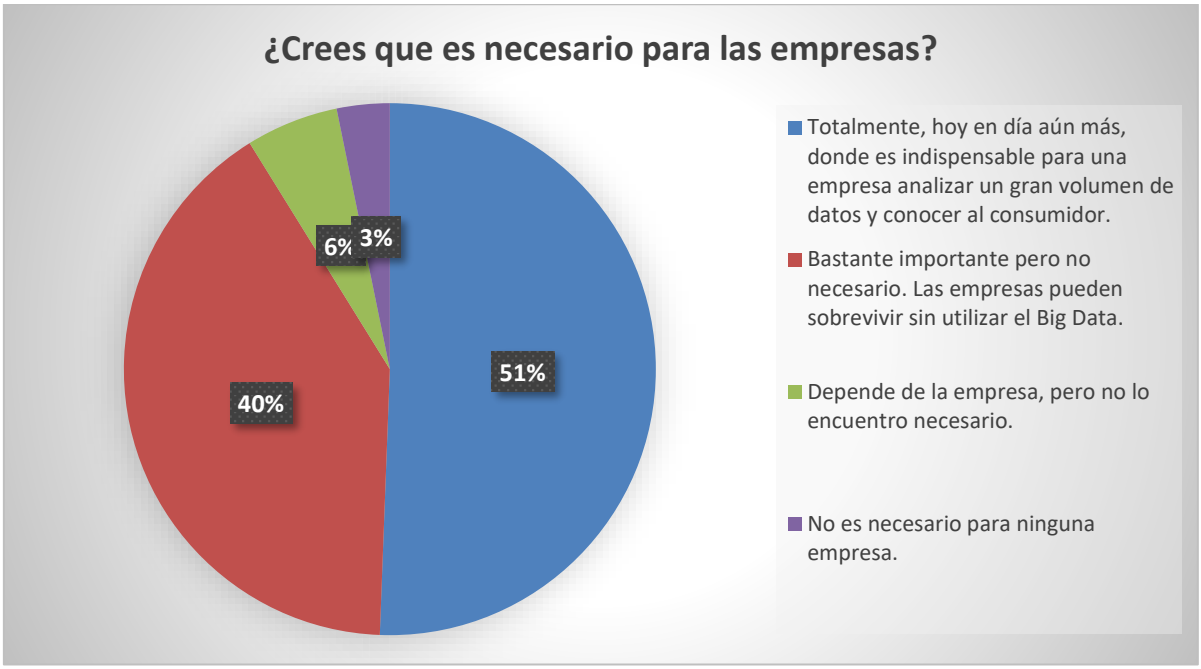
Gráfica 6: *Familiaridad de los encuestados con el término Big Data*. **Fuente:** Formularios Google

El 79,1% de los encuestados consideran que el Big Data es, sin duda, una gran ventaja competitiva para las empresas, y su utilización e implantación es muy importante. Un 12% no cree que suponga una ventaja competitiva pero sí proporciona un gran valor. Un 4,8% piensa que es una pérdida de tiempo para las empresas invertir en Big Data y un 4% considera que no aporta anda de valor.



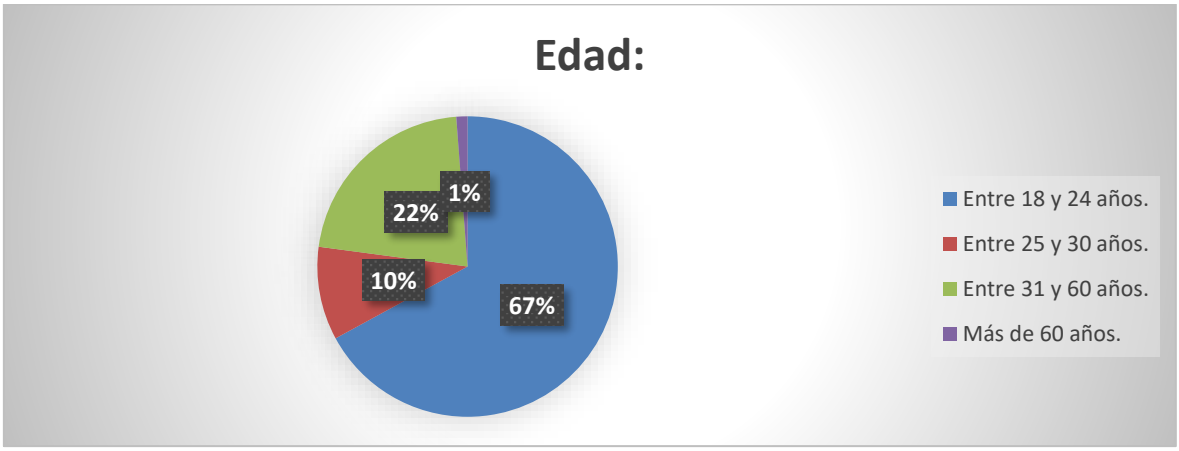
Gráfica 7: *Opinión sobre el valor que el Big Data proporciona a las empresas.* **Fuente:** Formularios Google

Es curioso que más de un 90% de los encuestados piensen que es totalmente necesario o bastante importante para las empresas. De este porcentaje, un 40,6% piensa que es importante pero que todas las empresas pueden continuar con su actividad sin utilizarlo. Un 3,2% piensa que no es necesario para ninguna empresa.

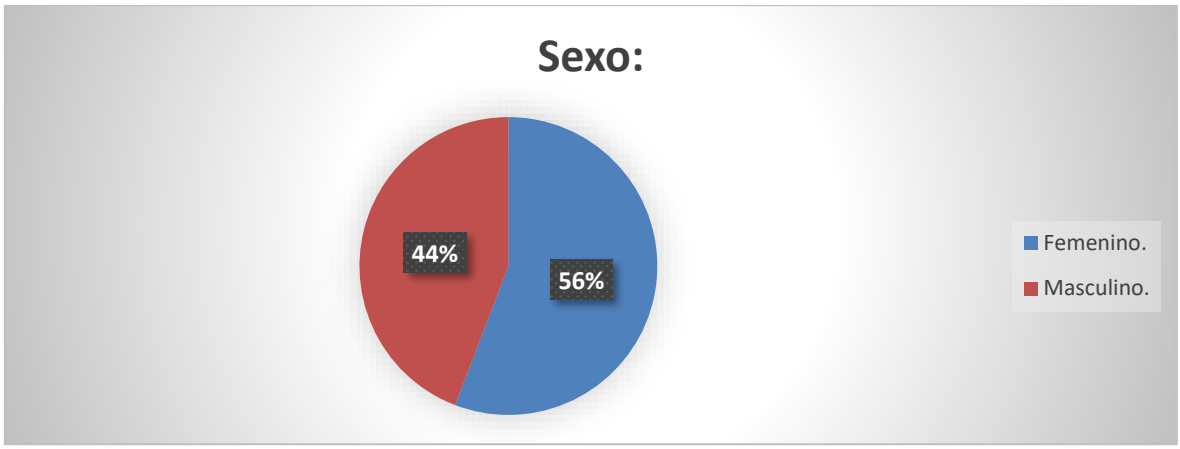


Gráfica 8: *Opinión sobre la necesidad de las empresas de implantar sistemas de Big Data*. **Fuente:** Formularios Google

Para terminar la encuesta, se preguntó la edad y el sexo de los encuestados, para ver si este aspecto puede influir en los resultados obtenidos. Un 67,1% de quienes han respondido a la encuesta tiene entre 18 y 24 años. Un 23% de los encuestados tienen más de 31 años. En cuando al sexo de los encuestados, se determinó que el 55,8% son mujeres y el 44,2% fueron hombres.



Gráfica 9: *Rango de edad de los encuestados.* **Fuente:** Formularios Google



Gráfica 10: *Sexo de los encuestados*. **Fuente:** Formularios Google

7 ENTREVISTAS A PROFESIONALES

Para completar el estudio de campo, además de conocer la opinión de los consumidores, se ha realizado dos entrevistas a profesionales del Big Data con el objetivo de conocer su visión dentro de una empresa, al igual que su visión de futuro. La finalidad es analizar la situación del Big Data dentro de las empresas contrastándolo con la opinión de los consumidores, para ver cómo puede afectarles o influenciarles a la hora de consumir.

De esta manera se conoce de forma más directa cómo se utiliza el Big Data, sus aplicaciones previamente explicadas en los primeros apartados del trabajo y la opinión más personal de un profesional que forma parte cada día del proceso de tratamiento de datos de forma masiva. Se buscaba por otra parte conocer el grado de información que tienen las pymes sobre Big Data, ya que al ser algo nuevo, aún existe mucho desconocimiento sobre sus aplicaciones, sus beneficios y su adaptabilidad al negocio que lleven a cabo.

Los profesionales que se ha entrevistado son los siguientes:

- Miguel Pírez Bustamante[20]. Trabaja a distancia con equipos multidisciplinares y participa en proyectos internacionales para la normativa europea bancaria. Especializado en la ingesta de datos con tecnologías Big Data y ETL.
- Diego Calvo Barreno[21]: Data Scientist que lleva a cabo proyectos Big Data para solucionar problemas en las empresas, mejorando su toma de decisiones y optimizando costes.

Trabajo de campo

Para llevar a cabo estas entrevistas, primeramente, investigué en LinkedIn en busca de perfiles de profesionales que pudieran aportarme información útil y acorde a mi trabajo de investigación. Quería centrarme en profesionales que trabajasen en Valladolid para poder contrastar posteriormente la información obtenida en las entrevistas con la plasmada en la encuesta realizada a los consumidores vallisoletanos. Pude contactar con dos profesionales que se dedican desde hace bastantes años al Big Data. Cada uno de ellos está especializado en una parte diferente del procesamiento de datos masivos y tienen experiencias diferentes, por lo que me podían aportar diferentes visiones y puntos de vista,

[20] Disponible en la URL: https://www.linkedin.com/in/miguel-pírez-bustamante-721352131/

[21] Disponibles en las URLs https://www.linkedin.com/in/diego-calvo/ y http://www.diegocalvo.es/big-data/

como ve puede ver posteriormente también en sus entrevistas. Me puse en contacto con ellos y pude acordar una entrevista. Ésta se llevó a cabo a través de Skype, donde podía tener una conversación cara a cara con ellos mientras grababa sus respuestas para analizarlas y transcribirlas posteriormente. Realicé una transcripción de éstas para poder conservar de manera precisa todas las palabras e información que me transmitieron. Las preguntas y el orden en el que se llevaron a cabo está pensado para ir contestando a cada uno de los apartados que se analizaron y explicaron previamente en el trabajo.

A lo largo de la entrevista se abordaron los siguientes temas:

- Aplicaciones que utilizan en su día a día para el procesamiento de datos.
- El Big Data en España.
- El valor del Big Data para las empresas, tanto para grandes como para pymes.
- Beneficios que obtiene el consumidor.
- Dificultades al tratar con los datos siguiendo la ley de protección de datos actual.
- El futuro del Big Data.

Las entrevistas completas se encuentran en el anexo junto con todas las preguntas realizadas.

8 CONCLUSIONES DEL ANÁLISIS

Tras haber realizado un análisis de la opinión de los consumidores a través de la encuesta previamente analizada, y de conocer la opinión de profesionales que trabajan en su día a día con tecnologías Big Data, he llegado a las siguientes conclusiones, que procedo a explicarlas a continuación:

- El Big Data está en auge y es una realidad. Cada vez son más las empresas que adoptan tecnologías para el tratamiento de datos de forma masiva, obteniendo grandes beneficios. Y aunque sea un concepto que todavía no esté interiorizado ni comprendido por los directivos de las empresas, poco a poco se va implantando.
- Llevar a cabo proyectos Big Data supone un gran coste para la empresa, ya no solo por las tecnologías que se deben implantar sino por el personal cualificado que se necesita. Por ello, muchas empresas han empezado a invertir en este tipo de proyectos grandes cantidades de dinero. En algunos casos, esta inversión no ha valido la pena, ya sea porque el proyecto no se ha podido llevar a cabo debido a la falta de personal cualificado que realmente conoce las herramientas, o bien porque no se han empleado las buenas tecnologías para el negocio. No en todos los casos se debería utilizar tecnologías Big Data, de hecho, por menos de 10 terabytes de datos no merecería la pena. Pero sí existen pequeños proyectos o adaptaciones de Business Intelligence que pueden ayudar mucho, por ejemplo, a los pequeños negocios.
- A la mayoría de las personas les importa que las empresas utilicen sus datos e información para su beneficio económico. Aunque solamente un 34% está muy preocupado por el tratamiento y posesión de su información por parte de las empresas, éstas tienen dos formas de excusarse. Según Google, la información es el precio que debe pagar el cliente a cambio de buenos servicios y aplicaciones por un precio muy reducido o simplemente gratuitamente. Según la filosofía de Telefónica, el consumidor tiene el poder de ceder o no su información, dependiendo de los servicios que quiera obtener.
- Cada año la ley de protección de datos es reforzada, sobre todo en el ámbito europeo. Lógicamente para las empresas, esto supone invertir muchos más recursos en llevar a cabo de forma correcta y legal todos sus procedimientos a la hora de tratar con datos personales. Es importante que los gobiernos estén a la altura y sepan adaptar la normativa a esta nueva era y revolución.

Por la mentalidad e ideas de la Unión Europea, se espera que el futuro consista en llevar a cabo mayores restricciones en el ámbito Big Data, dejando el completo poder de su información al cliente, siendo éste el que debe elegir si quiere ceder sus datos o no. Esta filosofía corresponde por la anunciada por empresas como Telefónica.

- Es importante que las empresas cumplan con la ley de tratamiento de datos personales, ya que hay un 30% de la población que dejaría de consumir los productos o servicios de una empresa si ésta utiliza su información personal de manera no autorizada o ilegal.
- Se puede obtener información de absolutamente todo, sin embargo, el teléfono y los CRM o trazabilidad de las empresas son las fuentes de recogida de datos más empleadas.
- Solo un 17% de los encuestados conoce las SmartCities o Ciudades Inteligentes como una forma de recogida de datos e información, esto puede deberse a que, en la mayoría de los casos, tan siquiera se conoce su significado.
- La mayoría de las empresas que utilizan estas tecnologías tienen una fuerte presencia en bolsa, esto incluye principalmente a las grandes empresas y a los bancos. Muchos de ellos desarrollan sus propios soportes y tecnologías, como Amazon y Google que poseen sus propios clústeres. Para ver la importancia y el beneficio que pueden obtener algunas empresas, siete de las diez primeras empresas con mayor cotización en bolsa se dedican a comercializar datos, como son Telefónica o Google a nivel mundial. Ambas tienen un servicio de venta de datos, de hecho, es un negocio al alza y que tiene un gran valor.
- En la población, cuanto mayor es la edad, existe un mayor desconocimiento sobre el procesamiento de datos. La mayoría no le da gran importancia, ni piensan constantemente en toda la información que están dejando al navegar, por ejemplo, en Internet. De hecho, muchos no son conscientes de la enormidad de fuentes de recogida de datos que existen y que las empresas utilizan a su favor.
- En base al punto anterior, un 88% de los encuestados han escuchado hablar alguna vez del Big Data, sin embargo, aún existe mucho desconocimiento. Muchos conocen el término, pero no saben cómo aplicarlo ni qué beneficios puede conllevar exactamente. De hecho, es raro el día que no aparezca en el periódico una noticia sobre ello.

- Un 9% de los entrevistados piensa que el Big Data es inútil, no merece la pena invertir en ello ya que las empresas no obtienen una ventaja competitiva. Sin embargo, los profesionales reconocen que, aunque hay que saber en qué invertir y qué tipo de proyectos llevar a cabo dependiendo de la empresa, es una herramienta muy útil y que proporciona grandes beneficios.
- El conocimiento sobre Big Data es mayor en la población más joven, es un tema de actualidad. La mayoría de los encuestados se encuentran en un rango de edad entre los 18 y los 24 años, y se ha visto, que a partir de esta edad el desconocimiento es mayor.
- El Big Data ha llegado a España, es un hecho, sin embargo, aún estamos más atrasados que en otros países, sobre todo que EEUU. Ellos son los pioneros, desde hace mucho tiempo que se lleva utilizando tecnologías y estrategias Big Data en el deporte. Eso ya ha llegado a España, las empresas se han dado cuenta, saben que está ahí y que hay que empezar a explotarlo.
- La ubicación de nuestro teléfono, así como todos nuestros movimientos con el Smartphone es lo que proporciona mayor información a empresas como Google. Y aunque con la nueva ley de protección de datos no puedan conocer el nombre de la persona, si bien puede reconocer muy fácilmente información lógica y esencial de los movimientos diarios de las personas.
- A partir del Big Data se ha generado otro nuevo negocio que es la compra-venta de datos o información. De hecho, muchas empresas no analizan ellos mismos información de forma masiva, sino que les es mucho más rentable comprar esta información debido por ejemplo a su tamaño. De la forma inversa, muchas de las empresas que más cotizan en bolsa actualmente se dedican a la venta de datos, como es el caso de la empresa Telefónica.
- El Big Data es una realidad, cada vez las empresas nos conocen más y eso es un hecho, la información es poder. Aunque va evolucionando constantemente, es algo que siempre va a estar presente, ya sea con nuevas tecnologías o con otro nombre. Es importante que los gobiernos busquen formas de informar más a los ciudadanos, ya que como se puede ver en los resultados obtenidos de la encuesta, a la mayoría les suena el concepto, pero no tienen conocimiento sobre ello. Y aunque ahora por ley, el consumidor debe ser informado en las cookies, igual el gobierno debería promover más acciones de este tipo.

- Según los profesionales del Big Data, las empresas son conscientes de que tienen a su disposición un gran volumen de información y que tienen que analizar estos datos. Lo que no saben es hasta qué punto deben hacerlo de forma masiva como es el Big Data y qué beneficios exactos les puede traer.

9 BIBLIOGRAFIA

Acens, (2014). "Base de datos NoSQL. Qué son y qué nos podemos encontrar" 2014. Recuperado el 9/2/2019 de http://bit.ly/2I97Ao6

Aguilar, L. J. (2016). "Big Data, Análisis de grandes volúmenes de datos en organizaciones". Alfaomega Grupo Editor.

Apache Hadoop, (2019). "Apache Hadoop, 2019". Recuperado el 23/2/2019 de http://hadoop.apache.org

Araujo, A. (2016). "¿Qué es una base de datos NoSQL?" Publicado el 19 de abril de 2016. Recuperado el 92/2019 de http://bit.ly/2HPAbQo

Areces, F. R. (2014). "Big Data: de la investigación científica a la gestión empresarial". *Revista de Occidente*, (400), 120-123.

Barrando R, (2012). "¿Qué es el Big Data?" Publicado el 18 de junio de 2012. Recuperado el 23/2/2019 en https://ibm.co/2MgW6UT

Bit, (2015). "Qué es el Big Data y para qué sirve" Publicado el 2 de abril de 2015. Recuperado el 4/10/2018 de http://bit.ly/2K9aKLg

BOE, (2018). Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal. Publicado en el «BOE» núm. 298, de 14/12/1999.

Camargo et al. (2015). J. Camargo-Vega, J. Camargo-Ortega, y L. Joyanes-Aguilar. "Conociendo Big Data". *Facultad de Ingeniería*, *24*(38), 63-77.

Cloudera, (2013). "Apache Hadoop Ecosystem" 2013. Recuperado el 23/2/2019 de http://bit.ly/2XmVcYa

Conesa et al. (2010). J. Conesa y J. Curto. "Introducción al Business Intelligence". *Editorial UOC. Barcelona. Mayo*.

Debitoor (2019). "¿Qué es el Cloud Computing?" 2019. Recuperado el de http://bit.ly/2XhjiDl

Driver, (2019). "Apache 2019". Recuperado el 23/2/2019 de https://www.apache.org

Galimany Suriol, A. (2014). "La creación de valor en las empresas a través del Big Data". Recuperado el 5/03/2019 de http://cort.as/-JGO-

Garcete, A. (2012). "Base de Datos Orientado a Columnas". Universidad Católica "Nuestra Señora de la Asunción," Asunción, Paraguay.

Gartner, (2012). "Reporte de Gartner analiza ´Big Data` alrededor de tecnología de datos" Publicado el 19 de noviembre de 2012. Recuperado el 13/10/2018 de http://bit.ly/30UNqXo

Gil, E. (2016). "Big data, privacidad y protección de datos". *Madrid: Agencia Estatal Boletín Oficial del Estado*.

Guilarte, M. (2012). "El último estudio Digital Universe de IDC revela una brecha en Big Data" Publicado 12 de diciembre de 2012. Recuperado el 3/10/2018 de http://bit.ly/2Xi6uN4

Hopkins, B. (2011). "Beyond the Hype of Big Data" Publicado el 28 de octubre de 2011. Recuperado el 14/11/2018 de http://bit.ly/2EJnXXJ

IBM, (2012). "Qué es el Big Data" Publicado el 18 de junio de 2012. Recuperado el 13/10/2018 de https://ibm.co/2MgW6UT

IDC, (2018). "Conferencia Big Data y Analytics" Publicado el 17 de septiembre de 2018. Recuperado el 14/10/2018 de http://bit.ly/2JM8yKg

IEP, (2019). "¿Qué es el Big Data? Ventajas y desventajas" 2019. Recuperado el 23/3/2019 de http://bit.ly/2WztIRT

Iglesias A. (2018). "Estas nueve empresas españolas son líderes en Big Data a nivel mundial" Publicado el 19 de diciembre de 2018. Recuperado el 9/3/2019 de http://bit.ly/2W5VIrZ

iic, (2016). "Las 7 V del Big Data: Características más importantes" Publicado 28 junio 2016. Recuperado el 3/10/2018 de http://bit.ly/2WzoPrU

INE, (2017). "Empresas españolas que analizaron datos Big Data" 2017. Recuperado el 9/3/2019 de http://bit.ly/3117Dee

INE, (2018). "Valladolid: Población por municipios y sexo" 2018. Recuperado el 18/4/2019 de http://bit.ly/2Z2XY5i

Informática para tu negocio, (2016). "Algunos beneficios del Big Data para tu empresa" 2016. Recuperado el 9/3/2019 de http://bit.ly/30VQuCR

Inmon, W. H. (2007). "The evolution of integration". *Inmon Consulting Services*. Recuperado el 4/04/2019 de http://cort.as/-JGOd

Interxion, (2015). "El 90% de las empresas se beneficia del Big Data" Publicado el 4 de agosto de 2015. Recuperado el 23/3/2019 de http://bit.ly/30T8eyy

Interxion, (2017). "4 consejos para adaptarse al nuevo reglamente de la protección de datos (GDPR)" Publicado el 17 de mayo de 2017. Recuperado el 6/4/2019 de http://bit.ly/2IaENdj4

Interxion II, (2017). "Reglamento general de de protección de datos (II): Otras facetas a tener en cuenta" Publicado el 25 de mayo de 2017. Recuperado el 6/4/2019 de http://bit.ly/2W3DTty

Jara, J. (2012). "Big Data & web intelligence". Paraguay: Universidad Católica" Nuestra Señora de la Asunción.

Jimenez, (2019). "Big Data: Qué es y cómo funciona" Publicado 21 marzo de 2019. Revista Forbes. Recuperado el 3/10/2018 de http://bit.ly/30YIQrt

Kyocera, (2017). "Diferencia entre datos estructurados y no estructurados" Publicado el 17 de marzo de 2017. Recuperado el 24/10/20218 de http://bit.ly/2HNA2wZ

López García, D. (2013). "Análisis de las posibilidades de uso de Big Data en las organizaciones". TFM del Máster de Empresa y Tecnologías de la Información y Comunicación. Universidad de Cantabria. Recuperado el 3/04/2019 de http://cort.as/-JGP4

Navarro (2017). "Big Data se hace fuerte" Publicado el 2 de noviembre de 2017. Recuperado el 24/10/20218 de http://bit.ly/2MgxBXX

Pérez, F. F (2019). "Tendencias tecnológicas y demográficas en ciudades. *contenidos*, 71".

Pérez, F. S. (2015). "Big Data. Economía industrial", ISSN 0422-2784, Nº 395, 2015 (Ejemplar dedicado a: Ciudades inteligentes), pp. 71-86.

Platzi, (2015). "¿Qué es el SQL y el NoSQL?" Publicado el 23 de febrero de 2015. Recuperado el 15/12/2018 de http://bit.ly/2HQrQw0

Prometeus Global Solutions, (2019). "Análisis del riesgo en la gestión de datos. RGPD y Big Data". Publicado el 22 de febrero de 2019. Recuperado el 30/3/2019 de http://bit.ly/2JQ1goX

Pulgar, et al. (2015). F. Pulgar-Rubio, C. Carmona, A. Rivera-Rivas, P. González y M. del Jesus**. "**Una primera aproximación al descubrimiento de subgrupos bajo el paradigma MapReduce". In *1er Workshop en Big Data y Análisis de Datos Escalable* (pp. 991-1000).

Puyol Moreno, J. (2014). Una aproximación a Big Data= An approach to Big Data.

ReporteDigital, (2016). "¿Por qué las empresas deben pensar en Big Data y cuáles son los beneficios que pueden obtener?" Publicado el 28 de noviembre del 2016. Recuperado el 23/3/2019 de http://bit.ly/2XkAGHw

Riquelme M. (2013). "¿Qué es un Data Mart?" Publicado el 2 de julio de 2013. Recuperado el 15/12/2018 de http://bit.ly/2EJgWG3

Rodrigues, (2012). "Emerging technologies for Big Data" 2012. Recuperado el 9/2/2019 de https://tek.io/2W8Pp6W

Sánchez, M. J. B. (2018). "Digitalización de los sectores turístico y financiero. Implicaciones jurídicas desde la perspectiva Big Data y fintech". *International journal of scientific management and tourism*, *4*(1), 217-235.

SAS, (2019). "Big Data; qué es y por qué es importante" 2019. Recuperado el 12/10/2018 de http://bit.ly/2JM6SQY

Serrano A., Gracia L., (2018). "Big Data y Protección de datos" 2018. Recuperado el 6/4/2019 de http://bit.ly/2Xn0frA

Tascón, M. (2013). "Introducción: Big data. Pasado, presente y futuro. *Telos: Cuadernos de comunicación e innovación"*, pp. 47-50. Recuperado el 21/04/2019 de http://cort.as/-JGQ3

That CS guy, (2018). "¿Qué es MapReduce?" Publicado el 20 de febrero de 2018. Recuperado el 9/2/2019 de http://bit.ly/2Wf3Pre

Urueña, et al., (2017). A. Urueña, M. Ballesteros y E. Prieto. "Informe e-Pyme 2016. Análisis sectorial de la implantación de las TIC en las empresas españolas". Recuperado el 23/04/2019 de http://bit.ly/2JWJS1V

Urueña, et al., (2018) A. Urueña, M. Ballesteros y E. Prieto. "Informe e-Pyme 2017. Análisis sectorial de la implantación de las TIC en las empresas españolas". Recuperado el 23/04/2019 de http://doi.org/10.30923/2341-4030-2018

Vidal, J. (2014). "Big Data. Gestión de datos no estructurados" Publicado el 27 de mayo de 2014. Recuperado el 24/10/20218 de http://bit.ly/2W1LDfq

ZDNet, (2010). "What is Big Data" Publicado el 16 de febrero de 2010. Recuperado el 14/11/2018 de https://zd.net/2QBrMme

10 ANEXOS

10.1 Entrevista a Miguel Pírez Bustamante

Hola buenas tardes Miguel, primero muchas gracias por haber aceptado a hacer esta entrevista. El objetivo de ella es conocer un poco mejor la visión y la utilidad del Big Data dentro de la empresa, en ese caso proporcionada por ti, que tienes experiencia trabajando con esta tecnología.

Me gustaría primero que te presentaras brevemente y que hablases un poco sobre tu trabajo actualmente y qué papel tiene el Big Data en éste.

Soy Miguel Pírez, yo estudié matemáticas. Actualmente lo que estoy haciendo no tiene nada que ver con las matemáticas. Aunque sí que hay un rol de matemáticos dentro del mundo del Big Data, en Valladolid en la empresa donde estoy no está. Podemos decir que la parte de matemáticas más orientada al Big Data es el analytics, es decir, realizar una interpretación de los resultados que se generan. Yo me dedico más al procesamiento de los datos, concretamente a una parte que dentro del mundo de la programación se llama ETL. Extracción, transformación y carga de los datos. En mi trabajo, yo me dedico a la parte de Big Data, la parte analytics no está tan desarrollada.

¿Con qué tecnologías de análisis Big Data trabajas en tu día a día?

Yo me dedico a la parte de ETL. Dentro del Big Data, hay una serie de perfiles, uno de ellos es el de programador. Dentro del programador hay varias áreas. Yo llevo trabajando dos años y medio en mi empresa, realmente he estado en un proyecto bastante grande para un gran banco. Ahí yo hacía una parte que es control de la calidad de los datos. Yo lo que hacía básicamente es que me pasaban una serie de datos y yo realizaba una serie de transformaciones o normalizaciones para que el banco después los tuviera limpios por así decirlo, para que tuvieran los datos que ellos querían. Que tuvieran un control de calidad. ¿Qué tecnologías utilizamos para eso? Yo lo que utilizaba es Hadoop. Eso es una palabra que dentro del mundo del Big Data es muy importante. El Big Data yo creo que se puede interpretar de dos maneras, una de ellos es el procesamiento masivo de datos, que no es algo nuevo, eso es importante, porque la gente piensa que los datos se han empezado a utilizar ahora. Pero no es algo nuevo, los datos los ha habido antes si nos remontamos al inicio de las computadoras, pero la diferencia es que cada vez hay más. Antes de hecho, el procesamiento de datos se llamaba Business Inteligente. Otra rama que

a lo mejor es la que se puede interpretar como Big Data es Hadoop, una tecnología nueva que surgió hace unos años que es de libre uso, y lo que permitió a través de unas nuevas tecnologías de procesamiento de datos un procesamiento masivo. Yo en mi trabajo lo que he estado haciendo en el último año, es con Hadoop, concretamente las herramientas de Hive, Spark, Scoop. Son herramientas derivadas de Hadoop, que Hadoop es lo que se llama un framework, que es un conjunto de herramientas que están metidas en un entorno. Yo básicamente me dedicaba al control de calidad del dato, o gobierno del dato.

¿Cómo crees que está la cultura y los conocimientos Big Data en España actualmente?

Bueno, creo que España es un país donde la tecnología ha tardado más en entrar si consideramos los países del primer mundo. Sí que es verdad que las empresas pioneras en aplicar Big Data son extranjeras, sobre todo en EEUU, que es donde empezó. EEUU ha sido el pionero, la vanguardia en tema de tecnología. En España pues sí, hoy en día, cada vez más empresas lo utilizan. Por ejemplo, yo trabaje con el banco Santander, más empresas se dedican a utilizar esos datos en su propio beneficio. Se han dado cuenta de lo importante y el valor que tiene utilizar esos datos. Otro tema que dentro de España ha tomado mucha vista respecto al resto, donde el Big Data ha tenido mucho efecto, ha sido en la liga, la liga española de fútbol, no es algo nuevo, el Big Data aplicado al deporte ha existido desde hace mucho en EEUU, como referencia te recomiendo una película que se llama Money ball, que habla como en EEUU ya hace muchos años, un ojeador de un equipo de baseball decidió utilizar las estadísticas y los datos para encontrar los mejores jugadores. Hoy en día uno de los casos en España que más claros hay y que la sociedad lo conoce es en la liga, donde esos datos valen mucho. En resumen, te diría, creo que efectivamente a España ha llegado el Big Data, es un hecho, es un sector en alce, las empresas cada vez se dan más cuenta de que es un valor añadido y es importante su desarrollo futuro. Es un recurso que han tenido ahí y que es necesario explotar.

En el análisis que estoy realizando, me he centrado sobre todo en encontrar un punto en común entre la visión del consumidor sobre la utilización de sus datos, y la visión de las empresas. He llegado a la conclusión de que una gran parte de la población no conoce muchas de las formas de obtención de datos. ¿De dónde se toman todos esos datos?

Las empresas hoy en día se han dado cuenta de que se generan datos, cada vez las empresas utilizan más los recursos electrónicos para organizarse. El número de datos que llegan es por infinitud de fuentes. Muchas empresas se han dado cuenta de eso y están empezando a hacer un datalake, es un repositorio de datos donde las empresas están a veces un poco obsesionadas por guardar todos los datos. Un datalake es un repositorio de datos en bruto. Muchas empresas dicen yo guardo mis datos, todos los datos que genere los guardo, y ya veré más adelante como puedo explotarlos. Eso por un lado, las empresas se han dado cuenta de que pueden guardar los datos y lo guardan en un datalake.

Por otro lado, ¿de dónde pueden obtener los datos? Los datos hoy en día los generamos de cualquier sitio. Hoy en día cualquier persona que tenga un smartphone, ese móvil genera datos por todas partes. Desde la ubicación hasta lo que buscamos. De hecho, la ubicación creo que es de las cosas más importantes que hay. La ubicación proporciona un número de datos increíbles para Google. A día de hoy, si no tienes implementados los servicios de Google no puedes hacer nada. Por curiosidad, esto lo leí en una entrevista por si lo quieres buscar, que se la hacen a un gurú del Big Data que se llama Martin Hilbert. Este hombre comenta que de las empresas que más cotizan en bolsa hoy en día, igual de las 10 primeras empresas, 7 se dedican a comercializar datos, como son Telefónica, o Google a nivel mundial. De hecho, Telefónica tiene su propio sistema de venta de datos. Por un lado, aparte de ofrecer un servicio a los clientes de conectividad a internet, llamadas… lo que es una telecom, por otro lado, también vende tus datos. De hecho, es un negocio al alza y que va a tener mucho valor en el futuro. Ya lo tiene y cada vez más. Hay muchas empresas, por ejemplo, Google, que tú dirías, cómo justifican éticamente la venta de esos datos. Pues Google piensa que a cambio de que la gente utilice mis aplicaciones y mis desarrollos informáticos gratuitamente yo me lo cobro utilizando sus datos. Tenemos que ser conscientes de que los datos están ahí, que tenemos que ser conscientes de que generamos datos y que esos datos valen dinero, y que las empresas que tienen acceso a esos datos como tu proveedor de servicios informáticos o el desarrollador de las aplicaciones que utilizas, los van a utilizar y los van vender. Es verdad, que hoy en día con la nueva ley, las empresas no tienen acceso a tu nombre en sí. Pero sí que es verdad que si una persona se mueve de una dirección a otra todos los días, eso significa que esa dirección son previsiblemente su trabajo y su casa. Que no saben tu nombre pero saben quién eres. Es un hecho, que hoy en día se puede saber muchas cosas sobre ti y que esas empresas van a utilizar esos datos y que valen mucho dinero y cada vez van a valer más.

¿Qué valor tiene estos datos para la empresa? ¿Qué beneficios le produce? ¿realmente merece la pena la inversión?

Si, hoy en día los datos son información y la información vale mucho. Si yo sé que cada vez más gente se mueve en bicicleta a su trabajo en las ciudades. Porque Google puede saber muy fácilmente en qué transporte te mueves. Entonces si ve que muchas personas están cogiendo desde tal sitio a tal sitio una ruta en bicis, pues una empresa que venda bicis puede decir pues voy a instalar en ese barrio, que por la razón que sea, está teniendo un auge en la venta de bicicletas, ya sea porque los servicios de autobús son muy malos, las vías de acceso o de ciclistas son mejores, puede haber muchísimos factores y por eso le interesa abrir una empresa ahí. O podría ser tan simple como una persona que vende panadería en el metro, dice pues mira yo sé que a esta hora, del metro sale más gente. Y puedes decir, esto es obvio, a las 3 o 7 de la tarde cuando las personas salen del trabajo pues hay más gente. Pero existen muchos factores a que a priori no son tan obvios que las empresas pueden utilizar. La empresa puede por un lado tanto procesar sus propios datos como comprarlos. La gente no piensa que sus datos se puedan comprar pero es un hecho, los datos se pueden comprar y se venden.

¿Crees que todas las empresas deberían incorporar tecnologías de análisis Big Data? ¿hasta qué punto también puede ser utilizado por pequeñas empresas o por negocios?

Todo esto te lo cuento desde mi visión. Sí que es verdad que todo lo que te he contado va más orientado de cara al cliente. Por ejemplo, una pequeña o mediana empresa, no solo tiene por qué querer tener datos de sus clientes. También puede dedicarse al IOT, internet of things. No se considera tanto del Big Data ero sí que es verdad que los datos que se consiguen con enormes. El IOT consiste, yo instalo en mi fábrica sensores por todas partes que me permitan tener un control en directo o casi en directo de todo lo que está pasando en mi fábrica. Pongo sensores a tal máquina que lo que hace es ver una prensadora y luego a los molinos que van moviendo las cacerolas, por ejemplo, si fabrico cacerolas de metal. No solo se tiene por que centrar en el tema de los clientes, hoy en día el Big Data se puede orientar al internet of things, los bancos lo pueden adoptar por ejemplo para calcular riesgos, es un tema que los bancos dedican cada vez más dinero y que es algo fundamental. También se mezcla con modelos matemáticos, pero a lo que voy es que no solo se orienta como te lo he enfocado hasta ahora, al tema de los clientes sino que dentro de la propia empresa también se puede automatizar determinadas partes del proceso de automatización de lo que quieras…no solo a los clientes.

¿Supone esto una gran inversión para las empresas?

Hombre, requiere inversión. Porque no es solamente crear el proceso y las infraestructuras para realizarlo. Hoy en día los gastos que conlleva viene de diferentes puntos. Por un lado, tienes que tener una estructura física, un clúster, que te permita guardar y procesar los datos, puedes montar tu propio clúster, que yo diría que solo lo realizan las grandes empresas, o bien puedes alquilar clúster a empresas como son Amazon; Amazon tiene sus propios clúster y los alquila; o como Microsoft con Azure, o diría que Google también tiene. Ahora no lo sé seguro, no sé cuál es el nombre de su servicio de alquiler, pero casi seguro que también lo alquila. No solo es desarrollar el correspondiente software que necesita para procesarlo sino también mantenerlo. Dentro de desarrollar el software hay que resaltar que también necesitas interpretarlo, por lo tanto, necesitas los profesionales de la parte de analytics que sepan interpretar esos datos y sacarles el provecho comercial.

¿Crees que la mayoría de las empresas conocen el valor que implantar estos sistemas de análisis de datos les puede proporcionar?

Yo creo que sí. No sé decirte. Yo no tengo una visión tan experimentada como para poder comparar con antes. El mercado del Big Data cada vez va a más. Ya existía antes como el Business Intelligence como te comentaba antes, pero hoy en día las empresas se dan cuenta de que tarde o temprano, si ya tenían un sistema como un sistema de datos más tradicionales, pues a lo mejor deciden que tienen que hacer una migración a una tecnología más puntera como Hadoop, que es una tecnología de software libre. Luego hay también empresas que se dedican a empaquetar ese software y dártelo directamente empaquetado. Pero en general yo creo que sí, no puedo decirte que a lo mejor las empresas españolas tanto, porque yo no me dedico a la venta de eso productos. Pero creo que sí, las empresas españolas se están dando cuenta. A lo mejor las pymes menos, porque es un desembolso no pequeño. Pero yo creo que sí, España está entrando en una revolución del Big data. Aunque no lo llamaría revolución, pero sí que es una revolución a más porque no es algo que recientemente haya surgido de repente como te comente antes, pero las empresas españolas están entrando y no quieren ir hacia atrás.

¿El consumidor gana algo con esto? Es decir, ellos están cediendo mucha información de forma gratuita la mayoría de las veces, o sin tan siquiera darse cuenta, pero realmente, ¿nosotros como consumidores qué obtenemos?

Cosas tan sencillas como un mejor servicio. Si nosotros, por ejemplo, hay una empresa, por ejemplo, el panadero que vende en una puerta del metro, si vamos a las 3 seguramente este hasta arriba. Si ese panadero sabe que a esa hora va a haber más clientes pues tendrá contratado a alguien que le ayude a esa hora. Debemos ser conscientes que siempre se habla de la parte negativa de si nos roban los datos, y es importante ser consciente de que los utilizan y que trabajan con ellos, pero también hay que ver la parte positiva. Desde luego la parte positiva puede ser por ejemplo en temas médicos, en medicina donde ya se está utilizando el Big Data y es claro el beneficio. Pero también es verdad que las empresas sepan Big data también puede ser bueno para nosotros si un banco te va a dar un préstamo, si realmente tú eres una persona que tiene poco riesgo de impago, es probable que por el Big Data te de sistemas más justos. No voy a entrar en eso porque cada banco tiene sus propios criterios. Pero sí que es posible que nos dé beneficio en el sentido de que nos dé una tasa de interés mucho más justa que con un cálculo tradicional de intereses nos podría dar. Y como te comentaba un mejor servicio. Sí que es verdad que vamos a estar más vigilados, pero también es verdad, que eso nos permite tener un mejor servicio de muchas cosas que antes no teníamos, como por ejemplo Blablacar, que es una empresa que se nutre de nuestros datos, que a través de internet y de saber quién quiere ir de un sitio a otro sitio, ha podido crear un sistema que a nosotros nos beneficia. Nos permite viajar más barato.

Yo creo que hay un beneficio mutuo. Yo creo que donde radica un poco el debate es en cómo puedo evitar que esto me pase. Si yo realmente no quiero que comercien con mis datos, ¿Cómo evitarlo? Eso ya es un tema más complicado, pero yo creo que claramente sí que hay un beneficio.

¿hasta qué punto las empresas tienen dificultades para trabajar con los datos de forma legal?

A ver para las empresas que haya una ley más fuerte sobre los datos personales de las personas es un incordio, significa poner más tiempo y dinero a controlar con esos datos que es lo que se hace. Si un banco por ejemplo utiliza datos personas de sus clientes y tiene que mandarlos a otras entidades ajenas en el extranjero. Por ejemplo, dentro de la UE no puedes mandar los mismos datos que fuera. Las empresas tienen que utilizar más recursos para tener un mayor control de los datos que están mandando y estoy de acuerdo

a la ley. Las empresas tienen que hacer un mayor esfuerzo. Si es necesario, pues sí, pero bueno, esto como todo pues ira evolucionando con el tiempo. Creo que es muy importante que los gobiernos estén a la altura y sean capaces de seguir el ritmo de cambio.

¿Hacia dónde va el futuro en relación al análisis de datos?

Creo que el Big data como tal no va a desaparecer nunca. El concepto de procesamiento y uso de datos, cada vez hay más y lo va a haber. La evolución de los últimos 10 años ha sido en producción de datos muchísimo mayor que los 10 años anteriores. Así que va a ir a más. Es posible que en el futuro surja un nuevo concepto, pero el concepto del procesamiento de datos va ir a más, es un hecho, no creo que nadie tenga una opinión contraria a esta. Es verdad que posiblemente luego cambie el concepto, porque es verdad que uno de los conceptos y particularidad del Big Data es una nueva tecnología que se llama Hadoop, es posible que luego llegue otra y llegará. Porque hace unos años las tecnologías tradicionales eran muy buenas y era el novamás y al final ha llegado una nueva tecnología para volverla obsoleta. Así que llegar otra nueva. Pero es importante que la gente conozca la realidad que hay. Que no es ninguna conspiración, que es una cosa que está cambiando y que va a seguir cambiando. Es bueno que la gente lo conozca y lo que estas intentando divulgar con tu tfg, que hay un negocio aquí. Los negocios son nuestros datos, y que lo tenemos que conocer. Que nada pasa por casualidad, que cuando entramos en internet y vemos algo no es por casualidad. Las empresas nos van a conocer más, es un hecho. Sí que es verdad que los gobiernos deberían buscar maneras no solo de controlarlo más sino también de informar al ciudadano más. Ahora las empresas te informan sobre las cookies, eso vino por una ley, igual los gobiernos también deberían seguir promoviendo acciones en este ámbito.

Mi visión es muy pequeña, yo de hecho no estudié informática ni nada, el Big Data me vino profesionalmente. Te recomiendo la entrevista de Martin Hilbert que te he dicho antes, es interesante, fue bastante sonado el caso de Cambidge Analitical por las campañas electorales estadounidense y cuenta como el Big Data se utilizó para las campañas electorales de Obama y Trump y cómo se utilizaron. Es un ejemplo bastante claro e interesante para la gente para saber cómo se utilizaron sus datos con un beneficio claro y luego tuvo unas consecuencias legales. Mark Zuckerberg también tuvo que pedir disculpas públicamente por un caso de filtración de datos también bastante sonado.

Hay negocio, lo hay, lo veo en mi día a día, las empresas cada vez demandan más y más. Y las empresas cada vez subcontratan a otras empresas para realizar proyectos. Me parece raro que la gente no lo conoce porque es rara la semana que no aparezca algo

sobre Big Data en el periódico. Llegará también a las universidades, los profesores de universidad no han oído hablar de eso. Pero llegará también, porque requiere personas de todos los campos.

Muchas gracias por tu tiempo, por tu opinión y por todo lo que nos has aportado.

10.2 Entrevista Diego Calvo

Hola buenas tardes Diego, primero muchas gracias por haber aceptado a hacer esta entrevista. El objetivo de ella es conocer un poco mejor la visión y la utilidad del Big Data dentro de la empresa, en ese caso proporcionada por ti, que tienes experiencia trabajando con esta tecnología.

Me gustaría primero que te presentaras brevemente y que hablases un poco sobre tu trabajo actualmente y qué papel tiene el Big Data en éste.

Yo te comento un poco de mi vida. Estoy haciendo el doctorado en temas de maxime learning, de redes neuronales y este tipo de cosas. Soy el technical lead del departamento de una empresa que se llama Bravent. He estudiado informática, también he estudiado investigación de mercados y también he estudiado la carrera de marketing, bueno, que se llama investigación de mercados, que es marketing. En cuanto a mi experiencia en temas de maxime learning y Big Data, llevo con estos temas unos cinco años más o menos. Y anteriormente tuve 8 años de experiencia en temas de dirección de proyectos, pero orientado a la parte de investigación y desarrollo. Eso es más o menos el contexto donde estoy moviéndome y lo que he hecho en mi vida.

Cuando empiezas en el mundo del Big Data, te das cuenta que no sabe casi nadie de qué va todo el tema del Big Data. Cuando vas a las empresas les suenan conceptos de Big Data y todo el mundo quiere subirse al carro, pero la mayoría de la gente no sabe de qué va. Todo el mundo quiere hacer proyectos y meter la parte de maxime learning. Pero como es guay, pues todo el mundo quiere entrar en el carro. Entonces yo lo que me he dado cuenta, desde mi perspectiva es que la gente quiere hacer proyectos, pero no sabe ni de qué va ni qué le va a aportar. Pero no a nivel técnico, te hablo a nivel de directores de empresa, de responsables de departamentos de tic, muchos no saben de qué va. Se están subiendo al carro muchas veces por el tema de que está de moda. Yo creo que no

se aprovechan todas las potencialidades del Big Data, primero porque no se conoce, no se conoce qué se puede hacer con ello.

¿Con qué tecnologías de análisis Big Data trabajas en tu día a día?

Tengo un blog sobre todo esto, que recibe unas 1500 visitas diarias, ahí puedes encontrar mucha información sobre las distintas tecnologías que utilizo. Yo ahora en mi día a día utilizo menos porque estoy en la parte de Microsoft y en la parte de Big Data hago menos cosas, pero anteriormente yo trabajaba para Telefónica con una serie de herramientas. Te lo voy a comentar en cuatro bloques, con el Big Data por un lado tienes la ingesta de datos, por otro lado, tienes el procesamiento de datos, tienes el almacenamiento de datos y luego la explotación o visualización de los datos. ¿Qué tecnologías he manejado en ingesta de datos? He trabajado con Flume, con Sqoop y con Nifi. Bueno y a mayores de los sistemas de ingesta de datos, hay sistemas de mensajería como Kafka y RabbitMQ o alguna de estas. Por la parte de procesamiento de datos he manejado Storm Streaming, una parte mínima de Flink. En la parte de almacenamiento de datos, lo que es la base de todo sistema Big Data es Hadsgs que es un sistema directivo distribuido que se base en partir los datos de una manera distribuida para que cada elemento tenga su parte ¿vale? Entonces yo he utilizado HDFS, y he utilizado también Hive, HBase y luego en telefónica utilice la parte de ElasticSearch. Y en la última parte, en la parte de visualización de datos, ahí está muy de moda Powervirai o Kivana. La cosa es que existen muchas tecnologías y muchas hacen prácticamente lo mismo. Cada compañía grande está sacando sus propias tecnologías, y cada parte del procesamiento está teniendo también sus tecnologías, de forma que el abanico de posibilidad para elegir es muy muy grande, dependiendo del tipo de proyecto se buscan unas cosas u otras.

Esto demuestra que cada parte del proceso Big Data, porque Big Data no es una sola tecnología, sino que es un contenido de tecnologías que nos permite hacer un procesamiento distribuido y que podamos atacar a grandes volúmenes de datos, tiene numerosas herramientas posibles, siempre dependiendo de cuál sea tu proyecto y tu finalidad, elegirás una u otra.

¿Cómo crees que está la cultura y los conocimientos Big Data en España actualmente?

Creo que hay mucho desconocimiento en toda esta parte. Creo que la gente se está poniendo las pilas muy rápido. La gente está haciendo cursos y tal, lo que pasa es que

están partiendo de una base de que nadie tenía ni idea. Entonces yo creo que está avanzando rápido la parte de conocimientos de Big Data, pero la oferta de Big Data sigue siendo mucho más alta que la posible demanda.

En el análisis que estoy realizando, me he centrado sobre todo en encontrar un punto en común entre la visión del consumidor sobre la utilización de sus datos, y la visión de las empresas. He llegado a la conclusión de que una gran parte de la población no conoce muchas de las formas de obtención de datos. ¿De dónde se toman todos esos datos?

Ahora se toma información de todo lo que pueda generar información. Desde un reloj inteligente pasando por un móvil, pasando por una red de sensores que pongas en el campo para la agricultura, a un chip que te implanten. Puedes encontrar cualquier cosa que emita un tipo de información y lo que se hace es guardarlo. De ahí se saca la mayoría, pero desde donde realmente se está extrayendo información históricamente son de los CRP o CRM de las empresas, y de toda la trazabilidad de sus sistemas de trabajo de las propias empresas. Pues, por ejemplo, el banco Santander, tiene un sistema de transacciones donde cada minuto va metiendo transacciones, pues a eso se le está poniendo una capa por encima y se está extrayendo información de ahí. Pero cada negocio lo extrae de donde lo tiene por así decirlo. Y a mayores están poniendo un montón de elementos más para poder generar más datos. Se están dando cuenta de que los datos es poder y es dinero y cada vez se están buscando más fuertes de información.

¿Qué valor tiene estos datos para la empresa? ¿Qué beneficios le produce? ¿realmente merece la pena la inversión?

Yo creo que hay que analizarlo, depende del sector y depende de todo. Ya te digo que hay gente que se mete a proyectos de estos grandes de un volumen de coste muy alto y si lo analizas sensatamente no merece la pena. Pero hay otras pequeñas inversiones en algunas empresas que merecen mucho la pena. Aquí lo que hay que irse es al coste-beneficio.

¿Crees que todas las empresas deberían incorporar tecnologías de análisis Big Data? ¿hasta qué punto también puede ser utilizado por pequeñas empresas o por negocios?

Siempre que me preguntan esto lo que yo les digo es que no pueden matar moscas a cañonazos. Es decir, la tecnología del Big Data es muy potente, pero hay otras tecnologías de Business Intelligence para analizar el negocio y extraer un conocimiento de ello que no necesitas Big Data. Estamos hablando de que por menos de 10 terabytes no tendría ningún sentido usar Big Data para mi forma de ver. Porque todo lo que te está aportando esa tecnología lo estás perdiendo porque tienes sistemas más complejos, gente más formada, costes más altos de máquinas. O sea, yo iría a ese orden de magnitud.

¿Supone esto una gran inversión para las empresas?

La parte gorda del gasto suelen ser las personas, porque estas tecnologías no se conocen entonces encontrar a alguien que controle estas tecnologías es caro y además tienen mucho riesgo, porque por salarios la gente se mueve mucho entonces te pueden dejar los proyectos tirados. Ahí tienes la parte del coste. Muchas veces prefieres a lo mejor utilizar tecnologías que te dan las cosas ya más hechas, aunque sean más caras y tengas que pagar mantenimiento, por el hecho de que prefieres algo más fácil para que haya otra persona que luego pueda continuar porque si se va la persona, te deja el proyecto tirado y fallido.

La inversión no es tan grande, depende de lo que tengas que hacer. Con más de cien mil euros puedes adaptar un pequeño proyecto que tenga tecnología de Big Data. Al final el mayor coste son las personas, si haces un pequeño proyecto con una persona, pues es el salario de una persona trabajando un año, más las máquinas que utilices, pues puedes ir a un proyecto de ese estilo.

¿Crees que la mayoría de las empresas conocen el valor que implantar estos sistemas de análisis de datos les puede proporcionar?

Yo creo que la parte de análisis de datos, lo que es extraer conocimiento de su negocio, sí que lo tienen en la cabeza. O sea, un responsable de departamento o el jefe de la empresa, sí sabe que eso se puede hacer y sí sabe que eso hay que hacerlo. Lo único que la parte del Big Data lo que te aporta es hacerlo por grandes volúmenes de datos, entonces ellos no saben ni son conscientes de si tienen que usar Big Data o no. Ellos saben que tienen que analizar datos y sacar conclusiones.

¿El consumidor gana algo con esto? Es decir, ellos están cediendo mucha información de forma gratuita la mayoría de las veces, o sin tan siquiera darse cuenta, pero realmente, ¿nosotros como consumidores qué obtenemos?

Pues mira vi el otro día una entrevista al director de telefónica, a Pallete. Por un lado, está la filosofía de Google, que dice que el consumidor obtiene servicios gratuitos a partir de aprovecharse de sus datos, y hay otra filosofía que es la que está implantando telefónica que es, los datos son tuyos, tú sabrás qué haces con ellos, si quieres venderlos eres libre de hacerlo, pero son tuyos. Entonces hay dos filosofías, la que, uno, tú tienes tu poder de los datos, y otra en la que las empresas supeditan el poder a darte servicios. Entonces, ¿qué gana el consumidor? Pues si tiras por la filosofía de Google, servicios gratuitos o a un mínimo coste, si tiras por la otra filosofía es que el propio usuario, por la propia tendencia del valor que tener datos tiene, entonces se puede cobrar por ello.

¿hasta qué punto las empresas tienen dificultades para trabajar con los datos de forma legal?

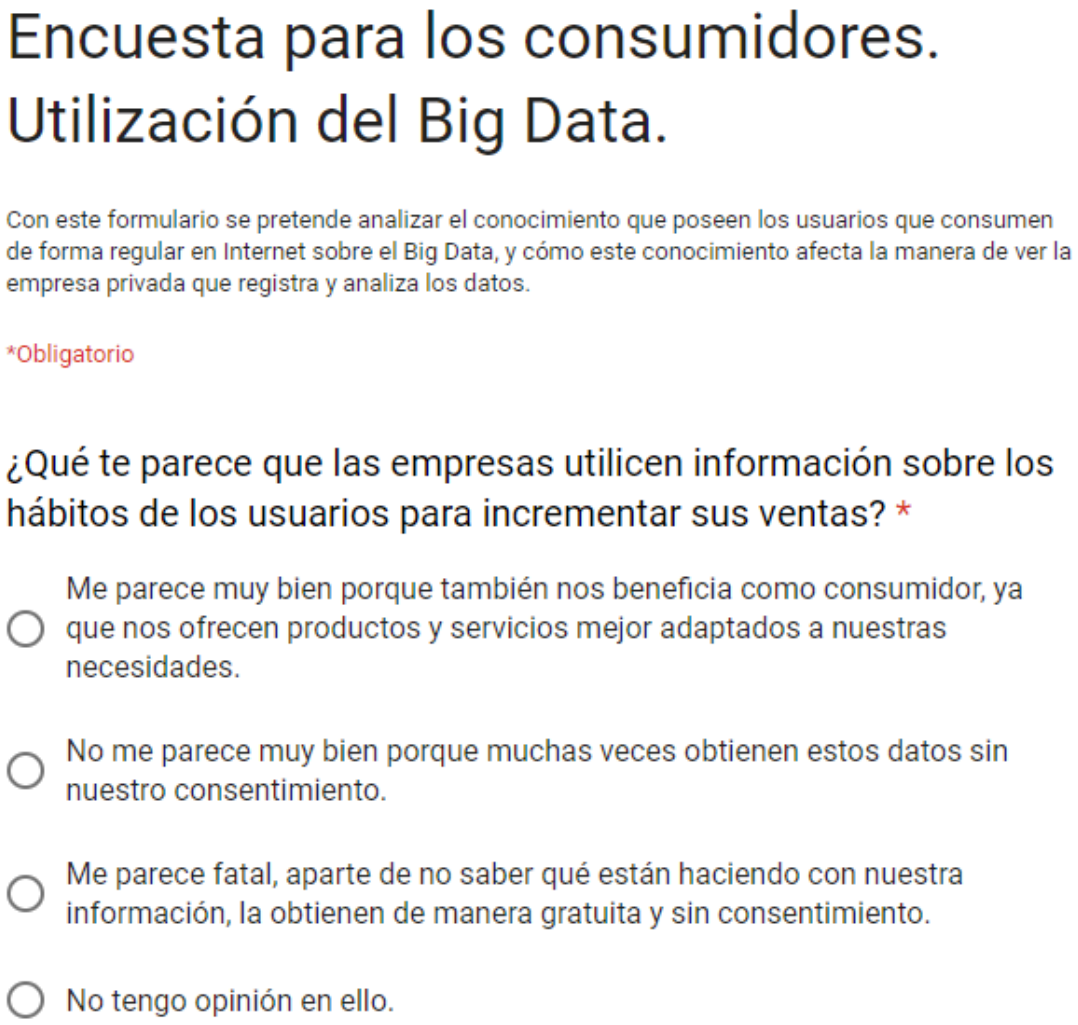
Antes la LPD era menos restrictiva, por una directiva europea hace poco cambió y empezó a atacar este tipo de prácticas para que el consumidor tuviese sus datos. Cuanta más legislación te ponga, más difícil es poder hacer lo que quieras con los datos. Mi opinión con lo que va a pasar, es que eso se va a tener que regular, no tiene ningún sentido no regularse, y yo creo que en un medio plazo, la Unión Europea, por los valores que tienen, van a ir más hacia una filosofía que te he comentado de Telefónica. Entonces yo creo que se irá regulando y que cada persona tendrá más poder de control de sus datos y de su privacidad. Al fin y al cabo, estás tentando contra la privacidad de la persona.

¿Hacia dónde va el futuro en relación al análisis de datos?

Yo veo al Big Data como una herramienta transversal que se va a meter en todos los sectores y va a hacer los sectores más productivos. Va a hacer que cosas que antes no se podían hacer como el coche autónomo o todo el tema como las SmartCities o todo este tipo de cosas que ya están sonando con bastante fuerza, van a ser posible gracias a estas tecnologías de Big Data. Está en auge y ha venido para quedarse. Es una moda, pero el volumen de datos cada vez va a aumentar más, se ha visto que da dinero, y en cuanto se demuestra eso, no va a bajar. Podría hacer la típica de no llegar a las expectativas y quedarse en su nicho de mercado, pero vamos, que ha venido para quedarse y lo tengo muy seguro.

Muchas gracias por tu tiempo, por tu opinión y por todo lo que nos has aportado.

Fragmento del cuestionario presentado a los consumidores a través de la plataforma de Formularios Google.



Encuesta para los consumidores. Utilización del Big Data.

Con este formulario se pretende analizar el conocimiento que poseen los usuarios que consumen de forma regular en Internet sobre el Big Data, y cómo este conocimiento afecta la manera de ver la empresa privada que registra y analiza los datos.

*Obligatorio

¿Qué te parece que las empresas utilicen información sobre los hábitos de los usuarios para incrementar sus ventas? *

- ○ Me parece muy bien porque también nos beneficia como consumidor, ya que nos ofrecen productos y servicios mejor adaptados a nuestras necesidades.
- ○ No me parece muy bien porque muchas veces obtienen estos datos sin nuestro consentimiento.
- ○ Me parece fatal, aparte de no saber qué están haciendo con nuestra información, la obtienen de manera gratuita y sin consentimiento.
- ○ No tengo opinión en ello.