

UNIVERSIDAD POLITÉCNICA DE CARTAGENA



**FACULTAD DE
CIENCIAS DE LA
E M P R E S A**

ANÁLISIS DE DATOS DE LAS ORGANIZACIONES. BIG DATA.

Juan José Liarte Muñoz

Directora: Soledad María Martínez María-Dolores

*Trabajo de fin de grado para la obtención del título de graduado en
Administración y Dirección de Empresa.*

ÍNDICE

Resumen.....	Pág 4
Introducción.....	Pág 4
1. DATOS.....	Pág 5
1.1. Tipos de datos.....	Pág 5
1.2. Fuentes de datos.....	Pág 6
2. BASE DE DATOS.....	Pág 9
2.1. Base de datos relacionales.....	Pág 9
2.2. Base de datos NoSQL.....	Pág 13
3. SOFTWARE DE CÓDIGO ABIERTO.....	Pág 17
4. ANALÍTICA DE DATOS.....	Pág 20
4.1. Analítica de datos en comercio minorista o retail.....	Pág 21
4.2. Analítica de datos y Big Data en la banca.....	Pág 22
4.3. Analítica web y social.....	Pág 24
5. BIG DATA.....	Pág 28
5.1. Importancia de Big Data.....	Pág 28
5.2. ¿Qué valor tiene para la empresa el Big Data?.....	Pág 30
5.3. ¿Cómo crear valor para el cliente con el Big Data?.....	Pág 34
5.4. ¿Cómo afecta el Big Data a la sociedad?.....	Pág 35
5.5. La importancia del Big Data en los negocios.....	Pág 36
6. Conclusiones.....	Pág 40
7. Bibliografía.....	Pág 42
8. Bibliografía de páginas web.....	Pág 43

ÍNDICE DE TABLAS

Tabla 1: Tipos de base de datos NoSQL.....	Pág 16
Tabla 2: Diferencias entre tipo de base de datos NoSQL Y SQL.....	Pág 17

ÍNDICE DE FIGURAS

Figura 1: Base de datos relacionales, un modelo detallado.....	Pág 10
Figura 2: Fila detallada de una base de datos relacional.....	Pág10
Figura 3: Muestra de una tabla basada en el sistema relacional.....	Pág 11
Figura 4: Ejemplo de informe de Google Analytics.....	Pág 26
Figura 5: Ejemplo de un segmento avanzado de Google Analytics.....	Pág 27
Figura 6: Ejemplo de un filtro de perfil de Google Analytics.....	Pág 28
Figura 7: El mercado de Big Data en el mundo.....	Pág 32
Figura 8: Sectores y tecnologías que impulsarán la inversión.....	Pág 33
Figura 9: ¿Qué quieren conseguir las empresas españolas con el Big Data?.....	Pág 34
Figura 10: Cubo OLAP.....	Pág 38

RESUMEN

El Big Data se está transformando en una de las inversiones más importantes para la empresa ya que nos permite obtener grandes ventajas para hacerla más competitiva. En este trabajo veremos la importancia del Big Data desde el punto de vista en la empresa y del cliente, para ello comentaremos los distintos tipos de datos que existen.

Podremos apreciar a lo largo de este trabajo la repercusión del Big Data en diversos sectores, además veremos los distintos tipos de base de datos, la importancia de la analítica web y social, para analizar dichos datos comentaremos algunos de los diversos softwares de código abierto que existen en la actualidad.

INTRODUCCIÓN

Antes de empezar, aclararemos el porqué de la elección de este tema. Como usuario de redes sociales y por el hecho de que las empresas tienen nuestros datos, nos preocupa la gestión que los mismos pueden hacer de ellos, por esa razón nos propusimos plasmar en el trabajo cómo lo gestionan las empresas. Lo primero con lo que nos encontramos es el término Big Data, una primera definición según Joyanes (2014) es: “son grandes volúmenes de datos con la necesidad de ser capturados, almacenados y ser analizados”. Por otro lado, Adrian Merv, vicepresidente de la consultora Gartner (2011) señala sobre este término que: “Big Data excede el alcance de los entornos de hardware de uso común y herramientas de software para capturar, gestionar y procesar los datos dentro de un tiempo transcurrido tolerable para su población de usuarios”.

Otra bastante significativa es la que realiza el McKinsey Global Institute, que en un informe muy reconocido y bastante referenciado, en mayo de 2011, define el término de otro modo: “Big Data se refiere a los conjuntos de datos cuyo tamaño está más allá de las capacidades de las herramientas típicas de software de base de datos”, esta explicación es, según ellos, subjetiva por lo que introducen otra condición refiriéndose a la magnitud que necesita un conjunto de datos para ser considerado Big Data. Destacan que ese volumen e incluso la definición en sí del término puede variar para cada sector, dependiendo de cuáles sean las herramientas de software disponibles para medirlo.

Se manifestó una convergencia de una abundancia de tecnológicas entre los años 2011 y 2013, y que se ha introducido con gran fuerza en empresas, en la sociedad, en los medios sociales y en los datos almacenados en la nube. (Joyanes L, 2014).

Estos datos han ido incrementando con el paso de los años. Por ejemplo, en el año 2012 se establecieron 1,8 zeta bytes de datos (un billón de gigabytes) según IDC. Todo ello quiere decir que el Big Data está emergiendo en todas partes y si los sabemos utilizar pueden proporcionarnos grandes ventajas de carácter competitivo para la organización o la empresa. Pero hay que tener también cuidado, ya que las empresas a su vez están expuestas a nuevos riesgos. Para ser competitivos en el mundo actual, como indica Frank (2012): “es imperativo que las organizaciones persigan agresivamente la captura y análisis de estas nuevas fuentes de datos para alcanzar los conocimientos y oportunidades que ellas ofrecen”.

Gartner cuantificó el desembolso en Big data en 2012 en veinte millones de dólares, pero no sólo eso, para 2013 alcanzó la cantidad de treinta y cuatro millones de dólares. Gartner no fue la única en realizar investigaciones acerca de Big Data, la compañía Deloitte 2012, estimó que más del 90% de las empresas pondrían en marcha iniciativas para el Big Data. (Joyanes L, 2014). Sin embargo, será en el año 2020 cuando un gran número de organizaciones trabajarán con Big Data, por consiguiente, se producirá uno de los mayores aumentos de los llamados Data Scientist. (Big Data Internacinal Campus 2018). Un Data Scientist es definida como: “Una persona con habilidades serias en diversas disciplinas técnicas, como ciencias de la computación, analítica, matemáticas, generación de modelos y estadísticas. Además, debe de ser un buen comunicador que sea capaz de entender un problema de negocios, transformar ese problema en un plano analítico, ejecutar el plan y luego dar una solución de negocios” (Kaul A, 2012, CEO de AbsolutData, empresa de analítica e investigación en California).

Podemos decir que tanto los profesionales en Big Data, como los analistas de datos y científicos de datos deben aplicar muchos esfuerzos porque tienen trabajo que hacer, y es de suponer que será uno de los empleos más solicitados en los próximos años.

En este proyecto, además de introducirnos en qué son los datos, en las diversas tecnologías que aplican las empresas y las distintas herramientas de Big Data, estudiaremos las diferentes maneras en las que la una organización puede obtener un mayor rendimiento en su toma de decisiones. La finalidad de este trabajo es plasmar la importancia de este tema, porque están acumulando nuestros datos y hoy en día, fácilmente, pueden ser hackeados y mal utilizados los sistemas donde se almacenan los mismos.

1. DATOS

1.1. Tipos de datos.

Es frecuente separar los datos en dos tipos: estructurados y no estructurados. Con el tiempo surgieron diversas herramientas para manipular los inmensos volúmenes de datos, por consiguiente, originaron un nuevo modelo conocido como dato semiestructurado.

➤ Datos estructurados.

Casi todas las fuentes de datos son de este tipo, dichos datos poseen una estructura fija. En estas fuentes los datos vienen acompañados de un formato definido y explicado al detalle que conforma la base de datos tradicional. Podemos poner ejemplos genéricos conocidos como las bases de datos relacionales o las hojas de cálculo.

Todos estos datos son conocidos por la empresa, puesto que se encuentran en un formato muy concreto. Estos formatos facilitan el trabajo con dichos datos. A modo de ejemplo mencionamos: documento nacional de identidad, número de cuenta corriente, e incluso, fecha de nacimiento. (Joyanes L, 2014).

➤ Datos no estructurados.

La principal diferencia de los datos estructurados a los datos no estructurados es que, éstos últimos son datos que no tienen un esquema predefinido. Son almacenados en los registros como documentos u objetos. Algún ejemplo de dato no estructurado que, por consiguiente, no tiene datos fijos son: audio, vídeo, fotografías, documentos, mensajes de correo electrónico o incluso formularios. Aproximadamente, el 80% de la información no se establece en las bases de datos de las empresas, si no que más bien, se encuentran dispersos en toda la organización y se conocen como datos no estructurados. (Joyanes L, 2014).

Sin duda alguna, si hay algún tipo de datos difíciles de analizar son los datos no estructurados, gracias a esta complejidad, ha provocado el nacimiento de diversas herramientas para su manipulación, por ejemplo: Hadoop o la base de datos de NoSQL. (Joyanes L, 2014).

➤ Datos semiestructurados.

Los datos semiestructurados tienen un esquema que puede ser definido, sin embargo, no resulta fácil entenderlos. Para poder leer los datos semiestructurados se requiere del uso de reglas complejas que explican cómo tratar cada pieza de información una vez que ha sido procesada.

Uno de los ejemplos más conocidos de los datos semiestructurados son los registros web logs de las conexiones a internet. Dichas conexiones están compuestas de distintos fragmentos de información, por ejemplos los textos de etiquetas de lenguaje XML y HTML. (Joyanes L, 2014).

1.2. Fuentes de datos.

El Big Data se obtiene de una cantidad abrumadora de diversas fuentes de datos. Pretendemos en este punto agruparlas de alguna forma para que nos hagamos una idea genérica sobre su procedencia.

A- Datos generados por máquinas (M2M)

Empezaremos explicando el significado de “machine to machine” o M2M. Estas siglas hacen referencia al intercambio de información en forma de datos que se lleva a cabo por dos máquinas conectadas entre sí. Todo sistema M2M cuenta con los siguientes elementos: máquinas que se encargadas de gestionar la información, un servidor que se encarga de gestionar el envío y la entrada de la información y una red de comunicación ya sea por cable, o bien, mediante señales inalámbricas. (Tendencias21, 2014).

Todos los elementos nombrados anteriormente hacen que se reduzca el tiempo y los costes y nos ofrecen un servicio que antes no disponíamos. Esta herramienta aumenta la productividad, aumenta la automatización y nos hace tener una vida cómoda y segura.

Esta portabilidad es debido a las tarjetas SIM; estas se pueden desmontar y van conectadas a los dispositivos móviles o tablets, permite dichos procesos y reducen la compleja instalación de estas máquinas, ayudando así a encontrar soluciones ante los diversos problemas que vayan surgiendo de forma más rápida. (Tendencias21, 2014).

Una de las empresas del sector de las telecomunicaciones, Movistar, creó una unidad de servicios M2M con más de cien profesionales de toda España que se dedicaron íntegramente a esta labor. El objetivo era ofrecer al cliente local y global una solución integral. Dicho programa consiguió reunir características muy avanzadas como, por ejemplo: detención de fraude, el control y la gestión del tráfico a través de datos informativos, incluso también, la capacidad de restringir comunicaciones entre ciertos dispositivos. (Tendencias21, 2014).

Este tipo de servicio tuvo un impacto directo. La iniciativa de la empresa surgió en Estados Unidos y desde hace tiempo es desarrollada en mercado español. Dicho proyecto fue liderado a manos de César Alierta, la estrella del mismo es conocida como “SmartSantander” con más de 20.000 dispositivos interconectados entre sí.

Actualmente, en España, casi tres millones de máquinas se comunican entre sí vía telefónica y se calcula que en 2020 llegará a los diez millones en todo el mundo, según Lucía Clemares (2014). El objetivo primordial entre las empresas consiste en mejorar los procesos en los sistemas de comunicaciones y para ello se van a centrar en diferentes escenarios: control de acceso, servicios de pago, servicios públicos y medio ambiente. (Tendencias21, 2014).

B- Internet de las cosas

“El IoT está formada por una cantidad impresionante de sistemas industriales interconectados, los cuales comunican y coordinan sus análisis de datos y acciones con un único fin, el de mejorar el rendimiento industrial. Su principal misión consiste en implantar redes de microcomputadoras que se conectan entre sí a través de internet, por ejemplo, un único pozo de petróleo que está dotado de 20 o 30 sensores puede generar hasta 500.000 datos cada 15 segundos. Se estima que la cifra será de 26 mil millones de dispositivos conectados a internet en el año 2020”. (Burns T, 2017).

Gracias a IoT, las empresas pueden sacar distintas ventajas. Destacamos de esta tecnología la posibilidad de conocer el comportamiento o los intereses de sus usuarios y pueden ofrecerles los productos o servicios que realmente necesitan. Otra de sus ventajas consiste en llevar un mejor control sobre el stock de los productos. Dependiendo del sector al que se desarrolle nuestra empresa podremos obtener también diversos beneficios. (Garatu, consultada en marzo 2019):

- Aquellas empresas que se dediquen al mantenimiento y control de carreteras, pasos a nivel o puentes, gracias al uso de IoT les permitirá llevar un control más exhaustivo de sus instalaciones.
- Las empresas que se dediquen al control atmosférico pueden acceder desde cualquier lugar a la información de sus sensores, para poder predecir movimientos sísmicos o determinar el clima atmosférico.

- Permitirá a las empresas llevar a cabo una mejor programación para la producción e incluso llevar un control en reparación y mantenimiento de sus instalaciones.
- Se podrá llevar una monitorización de nuestro consumo energético, y así conseguir una reducción del gasto energético, incluso en los hogares.

Sin duda, uno de los mayores beneficios de usar IoT para las empresas es la de gestionar grandes recursos, como, por ejemplo, la creación de ciudades inteligentes, lo que implica, la existencia de distintos dispositivos y sensores que permitan incrementar la calidad de vida de las personas, como por ejemplo a optimizar los recursos de energía y agua. (CIC, 2019).

A modo de ejemplo destacamos el proyecto Croke Park Stadium, llevado a cabo por la universidad de Dublín, Intel y por Microsoft, consistía en crear un estadio inteligente para innovar en IoT y estar cada vez más cerca de las ciudades inteligentes. Intel posicionó sensores por todo el estadio y de esta forma se recolectaban enormes cantidades de datos. A continuación, estos datos se guardaban en la nube de Microsoft Azure. Después, los investigadores de la universidad analizaban dichos datos para proporcionar al estadio una información eficiente. Gracias a la recogida de estos datos han podido: medir niveles de ruido, mantener el terreno de juego en perfecto estado e incluso monitorear la multitud para tomar diversas medidas de seguridad y salud. (Microsoft, 2016).

C- Datos sociales

Las redes sociales no son únicamente un canal de comunicación a nivel global, supone también una fuente de datos para las empresas. Por desgracia, la mayoría de ellos siguen sin saber cómo sacarles partido, pues sólo el 29% de las empresas son capaces de aprovechar esta información y de llevar ventajas respecto a sus competidores. En años anteriores, el 74% de las compañías empleaba datos recogidos en su página web y un 68% de empresas recogían y gestionaban datos a través de sus ventas. (Puromarketing, 2016).

Aun así, hay datos que no resultan tan agradables, puesto que hay compañías que hacen hincapié en que las estrategias de marketing no pueden beneficiarse de los datos sociales. (Puromarketing, 2016).

Unas de las formas más básicas y recomendadas para los usuarios es el monitoreo de las conversaciones, es decir, llevar chats con usuarios dentro de su página web y poder escuchar tanto las opiniones positivas como las negativas por parte de los usuarios para poder corregirlo en un futuro próximo. Este sistema es sencillo y recomendable, pero, no obstante, hay empresas que no lo tienen en consideración. (Puromarketing, 2016).

Sin embargo, hay organizaciones que sí cuentan con este sistema para monitorizar las conversaciones, de ellas, un 40% de las empresas creen que los datos son realmente procesables. (Puromarketing, 2016).

Para tener una ventaja competitiva y que nos distinga de nuestros competidores, debemos saber cómo gestionar los datos. En un análisis llevado a cabo por “SocialBro” (2016), se apunta que las empresas se enfrentan a dos problemas. Por un lado, está el problema de la métrica, es decir, cómo medir los datos recogidos. No resulta fácil de efectuar porque los sistemas tecnológicos y los chats no son capaces de leer los matices para poder separar, bien por ejemplo un elogio de una ironía y, por otro lado, están los datos introducidos por los usuarios. (Puromarketing, 2016).

Los datos sociales son necesarios y además son muy valiosos, ya que permiten una aproximación con el usuario para poder ofrecerle lo que necesita y poder escuchar sus opiniones, no hay que quedarse con los datos recogidos ya que no puede garantizar nada, hay que saber cómo gestionarlo porque gracias a ello podemos llevar una ventaja respecto a nuestros competidores. (Puromarketing, 2016).

2. BASE DE DATOS.

Para comenzar este apartado, explicaremos brevemente en qué consiste el término “bases de datos”. Se conoce como un grupo de información que se encuentra sistematizado de manera que su contenido pueda ser inspeccionado lo más rápido posible. Las bases de datos muestran datos estructurados con distintos parámetros, gracias a ello, el usuario puede encontrar lo que busca sin dificultad, además, si la base de datos estará perfectamente esquematizada facilita el acceso a la información actualizada y especificada. Sus componentes son: (Joyanes L. 2014)

- A. **Hardware:** se encuentra constituido por un dispositivo de almacenamiento, por ejemplo, un disco duro.
- B. **Software:** también llamado Sistema Administrador de Base de Datos.
- C. **Datos:** se encuentran almacenados según su estructura, después serán procesados para transformarse en información.

Existen diferentes tipos de bases de datos; hay bases de datos relacionales como MySQL, SQL y Oracle. Éstas hacen referencia a un modelo relacional y es recomendado utilizarlas cuando dichos datos son coherentes y la empresa tiene planeado utilizarlos.

También existen las no relacionales, como las NO-SQL (not only SQL). Este tipo es más flexible en cuanto a consistencia de datos y es la mejor opción para solventar alguna limitación o algún problema que puede llevar consigo el sistema relacional.

2.1. Base de datos relacionales.

La gestión de los datos por parte de las empresas ha estado principalmente gestionada por el sistema relacional. De todos ellos los más utilizados son: (Ionos, 2018)

- **Db2:** los usuarios disponen de un sistema capaz de gestionar las bases de datos relacionales cuya propiedad es de IBM.
- **MySQL:** consiste en una gestión de bases de datos de dominio público, además, de ser la más usada a nivel mundial.

- **Microsoft SQL server:** es un software desarrollado por Microsoft para llevar una buena gestión de las bases de datos, sólo es posible acceder a ella mediante una licencia de pago.
- **SQLite:** consiste en una especie de biblioteca, en la cual encontramos programas de código abierto que llevan incorporado un gestor de bases de datos relacionales.

El término de bases de datos relacionales hace referencia al sinónimo de “relación” propuesto por Edgar F. Clodd, (2013). Siguiendo al científico británico: “una relación representa un conjunto de entidades con las mismas propiedades. Cada relación se compone de una serie de filas o registros (llamadas tuplas) cuyos valores dependen de ciertos atributos (columnas)”.

Para establecer una relación o dominio, utilizamos el siguiente esquema con la siguiente sintaxis:

$$R = (A1:D1, A2:D2, \dots AN;DN)$$

La relación R hace referencia a los atributos A1, AN y cada atributo es relacionado con un tipo de dato, ya sea D1 y D2. Vamos a ver un ejemplo en la siguiente figura con el término de “empleado”.

Figura 1: Base de datos relacionales, un modelo detallado.

Empleados = (e_ID: integer,
 Primer apellido: string,
 Segundo apellido: string,
 Nombre: string,
 Número SS: string,
 Calle: string,
 CP: string,
 Municipio; string)

Fuente: Ionos 2018.

Como hemos visto, cada atributo se relaciona con un tipo especial de dato, esto quiere decir, que hay atributos que sólo aceptan una sucesión de caracteres (string) y otros que sólo admiten números (integer). Una vez que hemos visto el caso anterior, vamos a ver cómo sería en una fila en la siguiente figura.

Figura 2: Fila detallada de una base de datos relacional.

(1, Martínez, Fernández, Manuel, 46 12345678, Calle Segura 7, 30287, Cartagena)

Fuente: Ionos 2018.

La tabla es un formato que utiliza el sistema relacional para explicar visualmente el orden de los valores en una fila donde se muestran los atributos. A modo de ejemplo tendríamos la siguiente figura:

Figura 3: Muestra de una tabla basada el sistema relacional.

e_ID	1er apellido	2º apellido	Nombre	N. SS	Calle	CP	Municipio
1	García	Martínez	Ángel	45 23465438	Plaza Grande	30123	Cartagena
2	Martínez	García	Ramón	02 37458932	De América	30231	Cartagena
3	Segura	Fuente	Celia	12 47509935	Molina	30453	Cartagena
4	Muñoz	Revilla	Belén	78 341478945	Plaza Pequeña	30827	Cartagena

Fuente: Ionos, 2018.

Los datos estructurados en tablas corresponden a las bases de datos relacionales, el gestor de la base de datos se ocupa de definir su estructura y de gestionar los permisos necesarios para su modificación, cada uno de ellos tiene su lenguaje para realizar las siguientes operaciones: (Ionos, 2018).

- **Precisar el esquema de datos:** dentro de la definición de cada dato podemos hallar una breve explicación sobre la estructura del mismo, el cual, se encuentra almacenado en una especie de diccionario dentro sistema. Si, por ejemplo, un usuario, modifica o diseña una tabla, ésta se almacena en dicho diccionario con el esquema predeterminado.
- **Definir derechos:** los idiomas de las bases de datos deben proporcionar una sintaxis que ofrece la posibilidad de ceder o retirar permisos. A esto se le conoce como Data Control Language (DCL), vocabulario incorporado en la base de datos.
- **Definir condiciones de integridad:** hace referencia a las condiciones que se piden en un banco de datos, por ejemplo, que cada registro se identifique de manera indudable.
- **Definir transacciones:** en el momento en el que una base de datos pasa de un estado a otro distinto, estamos hablando de transacción. Dicha transacción debe seguir unas instrucciones que deben efectuarse en su totalidad. Cada transacción empieza con una orden para conectar a la base de datos, para después seguir otras que comienzan el procedimiento de datos, ya que también llevan consigo un paso de aprobación que asegura la integridad de la base de datos.

- **Definir vistas:** para generar una vista, el gestor de la base de datos genera una tabla virtual sobre la base de tablas. Dependiendo de la función de la misma se pueden apreciar distintas clases de vista, entre las más comunes son aquellas capaces de filtrar filas o columnas o también aquellas que conectan un número de tablas entre sí.

El lenguaje de bases de datos SQL se basa en el álgebra relacional. Las operaciones de las bases de datos como inventar, actualizar se llevan a cabo por sentencias SQL, esto quiere decir, que se tratan de una numeración de órdenes SQL relacionadas al inglés. Un ejemplo sencillo sobre una consulta a SQL sería: (Ionos, 2018).

```
SELECT columna FROM tabla WHERE columna = valor;
```

Utilizamos “SELECT” para mandar al gestor relacional una consulta sobre ciertos datos; después de eso, concretamos los datos que deseamos analizar facilitando la tabla y la columna que pretendemos escoger, y con “WHERE” incluimos un requisito en la sentencia SQL porque en nuestro ejemplo sólo queremos un valor determinado. Si volvemos a la tabla de empleados de la figura 3 la sentencia SQL quedaría así: (Ionos, 2018).

```
SELECT nº SS FROM Empleados WHERE e_ID = 3;
```

Con esto indicamos al gestor de la base de datos que muestre un valor de la columna nº SS de la tabla “empleados” y la restricción elegida fue la de seleccionar el valor del registro cuyo atributo de la columna e_ID sería igual a 3.

Finalmente, obtuvimos como resultado a Celia Segura Fuentes, con número de la seguridad social de 12 47509935 y que, lógicamente tiene el número 3 de identificación.

Las bases de datos relacionales basadas en SQL se caracterizan, sobre todo, en perseguir un planteamiento variable cuando organizamos los datos, entre las más comunes se encuentra las bases de datos orientadas a objetos y los sistemas basados en documentos. (Ionos, 2018).

Las bases de datos orientadas a objetos permiten almacenar los datos en forma de objetos, el resultado de todo ello consiste en productos con extensiones objeto relacional, éstos permiten facilitar el almacenamiento de los datos.

El almacenamiento en bases de datos orientadas a documentos se realiza asignando un determinado valor para cada clave. Dichos valores pueden ser cualquier tipo de información. Este sistema es la mejor alternativa cuando tratamos una gran cantidad de datos y además tiene la ventaja de que es un modelo adecuado si trabajamos con Big Data. (Ionos, 2018).

2.2. Base de datos NoSQL.

El término de NoSQL estuvo acompañado por la web 2.0, debido a que anteriormente las empresas que disponían de un portal podían subir contenido a la red, sin embargo, salieron a la luz distintas redes sociales como Facebook, Twitter o YouTube en el que cualquier usuario podría subir contenido, dando lugar a una creación masiva de datos. (Acens, 2014).

Debido a lo que acabamos de mencionar, empezaron a surgir los problemas para gestionar dicha información que se encontraba en bases de datos relacionales. Para poder solventarlo se dieron dos opciones posibles: en un principio las empresas se inclinaron por un elevado número de máquinas, aunque, desafortunadamente, esta opción no fue posible porque conllevaba un alto coste. La otra opción fue crear un sistema para un uso específico y con el paso del tiempo ha podido dar otras soluciones más robustas, dando lugar a los sistemas conocidos como NoSQL. (Acens, 2014).

Podemos decir que las bases de datos NoSQL son estructuras que nos ofrecen la posibilidad de almacenar la información en las circunstancias donde las bases de datos originan ciertos impedimentos, ya sean de escalabilidad o rendimientos de las bases de datos relacionales donde suben contenido millones de usuarios. (Acens, 2014).

Las bases de datos NoSQL se definen como sistemas en donde se guarda la información que no se formalizan con el modelo de entidad y relación, y tampoco usa un esquema de datos en forma de tabla, sino que emplea otro tipo de formatos como clave y valor, mapeo de columnas y gráficos.

A continuación, veremos las ventajas de este tipo de base de datos y su diferencia con el modelo que comentamos anteriormente, el modelo SQL.

Esta forma de almacenar la información presenta diversas características que lo hacen destacar sobre el modelo relacional. (Acens, 2014):

- **Se ejecuta en máquinas con pocos recursos:** estos sistemas de bases de datos se pueden ejecutar en máquinas con un importe mucho más moderado.
- **Escalabilidad horizontal:** una manera más fácil de mejorar estos sistemas es la de implementar más nodos.
- **Pueden manejar una gran cantidad de datos:** esto se debe a que utilizan un esquema distribuido.
- **No generan cuellos de botella:** una de los problemas que presentaba la base de datos SQL era que necesitaba transcribir cada tendencia para poder ser ejecutada y no solo eso, sino que cada sentencia requiere un nivel de ejecución todavía más complejo, lo que constituye una fuerte desventaja para los sistemas SQL.

Algunas diferencias destacables que nos podemos encontrar a la hora de usar de estos modelos de bases de datos son: (Acens, 2014):

- A. **No utilizan SQL como lenguaje de consultas:** casi todas las bases de datos de NoSQL emplean este idioma como soporte. En el caso de “Cassandra”, un tipo de base de datos NoSQL, usa el lenguaje CQL o, por ejemplo, BigTable que utiliza el GQL.
- B. **No utilizan estructuras como, por ejemplo, tablas para almacenar los datos:** los sistemas de bases de datos de NoSQL hace uso de otro tipo de modelo para almacenar la información como son los sistemas de clave/valor, objetos e incluso gráficos.
- C. **No suelen permitir operaciones JOIN:** como disponemos unas grandes cantidades de datos, los sistemas de bases de datos NoSQL prefieren eludir los JOIN. Definimos JOIN como un procedimiento que nos permite informarnos sobre diversas tablas de datos conjuntamente. (Ionos.es, 2018). Cuando la ejecución no se trata de la investigación de una clave, este exceso puede llevar consigo un alto coste, la solución más rápida para este inconveniente es la de realizar el JOIN mediante software.
- D. **Arquitectura distribuida:** normalmente las bases de datos relacionales están centradas en una sola máquina, pero en los casos de base de datos NoSQL la información se encuentra compartida en distintas máquinas gracias a unos mecanismos de tablas bien diseñadas.

A lo largo de los años se han creado diversas variedades de bases de datos para dar soporte a necesidades específicas o porque los antiguos modelos se quedaban obsoletos. Vamos a enumerar algunos de los tipos de base de datos NoSQL más importantes. (Tecnologiasinformacion, consultada en diciembre 2018)

- **Bases de datos de valores clave:** estas bases NoSQL de valor/clave destacan en por su sencillez y son bastantes eficientes para permitir a la aplicación la escritura y lectura de datos no transaccionales. Los datos que se almacenan pueden llegar a ser de cualquier tipo de objeto binario (texto, documento o vídeo). El programa controla todo lo que almacena, y ésta, a su vez, lo convierte en los modelos NoSQL más flexible. Una vez realizado ese paso, los datos se reparten y se replican en un clúster para poder obtener una característica esencial de este tipo de base datos, para poder ser escalable y estar disponible.
- **Base de datos de documentos:** estas bases de datos NoSQL son capaces de almacenar documentos JSON, XML y BSON. Se puede afirmar que son similares a la base de datos de valores clave, la única diferencia respecto al modelo anterior es que un valor es un solo documento capaz de almacenar todos los datos relacionados con una clave precisa.

- **Bases de datos de columnas anchas:** este tipo de base de datos NoSQL de columnas anchas tiene la capacidad de almacenar los datos en tablas con filas y columnas parecidas a las bases de datos relacionales, una característica que las diferencia de las bases de datos relacionales es que los nombres y los formatos de las columnas pueden ser modificados de fila en fila, también se diferencia en que una consulta puede recobrar datos relacionados con una operación, porque solo se recuperan las columnas asociadas con la consulta.
- **Bases de datos de gráficos:** este tipo de base de datos NoSQL utiliza los gráficos para almacenar, o consultar distintas relaciones. La información se presenta con forma de nodo de un gráfico y sus relaciones con las distintas aristas, de manera que sea fácilmente entendible la información que representa. Este modelo nos ofrece una navegación rápida y eficiente, a diferencia de los demás tipos de bases de datos NoSQL.

A continuación, presentaremos una breve explicación de algunos de las principales bases de datos NoSQL:

A- Cassandra

Cassandra nos proporciona la característica de distribución y gracias a ella la información se encuentra distribuida a lo largo de los nodos, por consiguiente, si alguno de los nodos se cae, el servicio no se degrada.

El rendimiento de Cassandra es lineal, lo que significa que, si tenemos 2 nodos podemos soportar 10.000 operaciones, ahora si disponemos de 4 nodos podremos soportar 20.000 operaciones. (Paradigmadigital, 2016).

B- Redis:

Esta base de datos NoSQL es la más conocida y por supuesto la más utilizada dentro del segmento de clave/valor. Lo que hace popular a Redis es su alta velocidad, ya que permite tener la información en memoria, además es conocida por su flexibilidad y su sencillez de uso. (Paradigmadigital, 2015).

C- MongoDB:

Este tipo de base de datos NoSQL está elaborada principalmente para documentos, es decir, que en lugar de guardar los datos en un registro se encarga de guardarlos como documentos, y éstos se almacenan en una herramienta conocida como BSON que se puede decir que es una representación binaria de JSON. Definimos a JSON como un formato de texto en el que nos permite permutar datos. (Paradigmadigital, 2015).

La diferencia primordial con respecto a las bases de datos relacionales es que no siguen un formato determinado, sino que cada documento puede llevar un esquema distinto. Este tipo de base de datos NoSQL viene por defecto con una consola en la cual podemos realizar distintas órdenes, además, está construido sobre JavaScript, quiere decir que las consultas se realizan usando ese lenguaje. (Genbeta, 2014).

D- CouchDB:

Este tipo de base de datos NoSQL es capaz de adaptarse tanto a ambientes de servidor como a clientes de diversos tipos. La característica de este tipo de base de datos trata de un sistema que permita que una variedad de productos pueda comunicarse entre sí, compartiendo datos entre ellos y llevando a cabo una sincronización de manera automática.





Existe una amplia variedad de tecnologías que usan el protocolo de replicación de CouchDB, por ejemplo, teléfonos móviles o navegadores, dicho protocolo permite que los datos de CouchDB puedan alojarse en varios sistemas. (Arsys, 2017).

E- Neo4j:

Esta base de datos NoSQL es capaz de representar datos y las relaciones entre ellos. De todos los que usan gráficos, Neo4j es la más conocida y también la más usada. Las bases de datos orientadas a gráficos responden a las consultas actualizando el nodo y las relaciones de esa búsqueda y no a todo el gráfico completo, lo que agiliza bastante el proceso. (Bbvaopen4u, 2015).

En la tabla 1 mostraremos un resumen de los tipos de base de datos NoSQL.

Tabla 1: Tipos de base de datos NoSQL.

Documentales	Datos semi estructurados en documentos (XML, JSON y BSON)	
Gráfico	Datos estructurados como nodos relacionados entre sí y representados a través de gráficos	
Clave/valor	Introduce datos estructurados como clave/valor	
Familia de columnas	Datos estructurados en diversas columnas, donde cada una de ellas puede tener un estructura distinta	

Fuente: Elaboración propia.

Una vez que hemos visto los tipos de base de datos tanto SQL como NoSQL, ¿Cuál deberíamos elegir para nuestra empresa? En la tabla 2, podemos ver las diferencias entre los distintos tipos de base de datos hasta ahora analizados.

Tabla 2: Diferencias entre tipos de base de datos NoSQL y SQL.

SQL	NoSQL
Permite la unión de tablas mediante JOIN.	No permite la unión de tablas con JOIN, y si lo incluye, está muy limitado.
Conlleva una dificultad para distribuir los datos.	Proporciona gran facilidad para distribuir los datos.
Presente una escalabilidad vertical, para mejorar la potencia del servidor y obtener mejores resultados.	Presenta una escalabilidad horizontal, es capaz de repartir bases de datos en diferentes servidores.
Tiene la habilidad de crear restricciones, prohibiciones y claves sobre los datos.	Estas restricciones o prohibiciones no están disponibles en NoSQL.
Los datos se rigen siguiendo unos esquemas.	Los datos no se rigen por esquemas.

Fuente: Elaboración propia.

Basándonos en el negocio de una tienda online, por ejemplo, la empresa DosFarma, escogeríamos el modelo de NoSQL, ya que esta base de datos es mucho más abierta y flexible ya que permite que nos adaptemos a las necesidades del proyecto de manera más sencilla, conlleva una escalabilidad horizontal, es decir, nos permite crecer en número de máquinas en lugar de invertir en máquinas de alto coste. Empresas tan conocidas como Amazon, EBay, El Corte Inglés o Carrefour, utilizan el modelo NoSQL, debido a que estas empresas también llevan a cabo el negocio online.

3.SOFTWARE DE CÓDIGO ABIERTO

La revista CIO (2012), publicó una lista de tecnologías de código abierto para Big Data. Dicha revista considera que los programas de código abierto son la principal iniciativa de Big Data, y por eso pasamos a hacer una breve relación de los mismos a continuación.

➤ Apache Hadoop.

Apache Hadoop se desarrolló pensando en la idea de un programa que permitiese tratar con enormes volúmenes de datos, ya sea por medio de clúster o bien mediante un patrón de programación. Proporciona un marco de trabajo que está escrito en Java, en el que permite crear nuevas aplicaciones que exijan un uso constante de datos. Este proyecto es dirigido por “Apache Software Foundation”.

“Los principales colaboradores de Hadoop trabajan en algunas de las compañías más grandes del mundo de la tecnología. Hadoop es utilizado en finanzas, telecomunicaciones, entretenimiento, gobierno e incluso en instituciones de investigación. Originalmente, esta plataforma se diseñó para analizar los datos estructurados y complejos, gracias a ello, muchas empresas optan por desplegar Hadoop junto a sus sistemas informáticos”. (Rodríguez F. 2014).

Hadoop es realmente útil cuando se va a llevar a cabo proyectos que necesite una estabilidad. Este software tiene origen en Apache Nutch, que es un motor de búsqueda de código abierto. En 2004, Google difundió un artículo en el que consiguió introducir MapReduce y en 2005 la mayoría de los algoritmos Nutch habían sido efectuados por MapReduce y NDFS (sistema de archivo de Nutch). (Rodríguez F. 2014).

Para enero de 2008, Hadoop formó su proyecto con un nivel superior a sus anteriores y obtuvieron un gran éxito. En aquella época, Hadoop estaba siendo utilizada por Yahoo!, Facebook o el New York Times.

Una vez que hemos comentado un poco de su historia, vamos a aclarar su estructura, (Rodríguez F. 2014):

- **Sistema de ficheros:** para que Hadoop pueda llevar un correcto funcionamiento utiliza un sistema de ficheros llamado HDFS.
- **Hadoop MapReduce:** el motor de Hadoop está formado por un planificador de trabajo MapReduce, y también está formado por una sucesión de nodos que se encargan de llevarlo a acabo.
- **Hadoop Common:** gracias a este apartado, el sistema puede sacar provecho por la integración de subproyectos de Hadoop.

Sobre el sistema de ficheros se encuentra el motor de MapReduce, el cual se fundamenta en un organizador de las tareas denominado JobTracker. Gracias a este planificador, el trabajo consistirá en ejecutar las funciones en cada nodo. Si el trabajo no puede ser ubicado en uno de los nodos, se les asigna prioridad a nodos en la misma consola, gracias a ello se reduce el tráfico de red y, si un nodo falla, esa parte del trabajo se replanifica. Por otro parte, en este sistema de ficheros se encuentran el NameNode y el DataNode. El primero de ellos se encuentra en el nodo maestro y su función consiste en decirle al sistema donde la ubicación de los datos que buscamos; los NameNode se encargan de mantener almacenados los datos. (Rodríguez F. 2014).

Tenemos que añadir que, Hadoop se puede efectuar por tres modos distintos, (Rodríguez F. 2014):

- **Modo local:** este modo de ejecución es el que viene por defecto, está configurado para una ejecución de forma no distribuida.
- **Modo pseudo-distribuido:** Hadoop ofrece la posibilidad de ser ejecutado en un modo distribuido, donde se ejecutan en procesos Javas diferentes.
- **Modo distribuido:** gracias a este modo se aprovecha toda la capacidad de Hadoop, consiste en poder aprovechar los recursos disponibles para maximizar los distintos procesos que lleva a cabo la plataforma.

Hadoop es una plataforma esencial para llevar a cabo nuevos retos y dar solución a los diversos inconvenientes o problemas que no se podían afrontar, gracias a su versatilidad, las empresas han encontrado usos diferentes para esta plataforma. Vamos a comentar diversos ejemplos prácticos del uso de Hadoop implementados para aprovechar las grandes oportunidades que ofrece Big Data, (Powerdata, 2015):

- **Ejemplos de tiempo real para la toma de decisiones:** control de fraudes, rendimiento de infraestructuras interconectadas, entender comportamientos del cliente.
- **Ejemplos en departamentos:** investigación y desarrollo (I + D), marketing y ventas, conocer el resultado de una campaña de marketing, beneficios de las empresas, conocer mejor al cliente usando información externa, personalizar promociones, e incluso valoración de la marca.
- En el mercado de la **banca** es común usarlo para mejorar ofertas, modernizar el servicio a los clientes o revelar posibles riesgos de fraude, mientras que, en el sector de los **seguros**, además de lo que hemos comentado anteriormente, se beneficia de la analítica del rendimiento del producto.
- En el sector de las **telecomunicaciones**, pueden ofrecer servicios basados en localización y también de muchas otras posibilidades de que vayan surgiendo según nuestras necesidades.
- El **retail** o comercio al por menor, se aprovecha de Hadoop para ganar en competitividad con analítica del cliente y fidelización del mismo.
- La optimización del comercio digital, el nivel de audiencia o la transformación del proceso de negocios son propios de sectores de **entretenimiento**, y en el **transporte y el turismo** que se benefician de la analítica del cliente y su fidelización o de la optimización de los precios.

Nos gustaría hacer una mención especial respecto a la gestión pública. Hadoop ha traído soluciones a diversos servicios tales como: servicios fiscales y tributarios e incluso a diversos temas relacionados con la defensa y la inteligencia.

- R.

Este software de código abierto está diseñado para la computación y visualización estadística. Se convirtió con gran rapidez en una herramienta esencial para analizar datos estadísticos de un gran conjunto de datos. Este lenguaje de programación está siendo comercializado por la empresa Revolution Analytics. Sin embargo, este software tiene dos desventajas: sólo puede manejar conjunto de datos del tamaño de la memoria de la máquina. y posee una menor escalabilidad. (Joyanes L, 2014).

- Jaspersoft.

Este software de código abierto es uno de los que tienen una mayor implantación en las empresas. Jaspersoft proporciona diversas herramientas para la inteligencia de los negocios. Este software nos da la posibilidad de acceder a fuentes de Big Data tales como base de datos NoSQL. (Joyanes L, 2014).

A modo de ejemplo destacamos la empresa SAS, una empresa líder en analítica de datos, mostró algunos casos de Big Data. El que más impactó fue el del equipo de baloncesto de la NBA, Orlando Magic. Este equipo de baloncesto colaboró con el Big Data para incrementar su rendimiento, vender más entradas y aumentar su eficacia como organización. Utilizó herramientas de SAS y de analítica de datos, todo ello les ayudó para situarse del duodécimo lugar al séptimo en la liga NBA. (Joyanes L, 2014).

Otro caso presentado por SAS, fue el de un banco africano, el cual comenzó a monitorizar las redes sociales y encontró que sus clientes no estaban satisfechos. Gracias a esto, realizaron nuevas ofertas a sus clientes a través de redes sociales y pudieron monitorizar su nivel de complacencia. (Joyanes L, 2014).

4. ANALÍTICA DE DATOS

A día de hoy, los datos son uno de los componentes principales de una empresa. Cada vez más las grandes organizaciones obtienen ingresos millonarios por la gestión de las bases de datos de sus usuarios. (Camaravalencia, 2015).

“La analítica de datos implica los procesos y actividades diseñadas para obtener y evaluar datos para extraer información útil”, Data analytics, ISACA (2011). Dichos resultados se pueden utilizar para: identificar áreas clave de riesgo, fraude, errores, para mejorar procesos de los negocios o incluso influir en las decisiones del negocio.

El concepto de analítica de datos da respuestas a preguntas tales como: ¿Cómo me puedo diferenciar de la competencia? ¿De qué forma se podría configurar el recorrido de los compradores en la tienda? ¿Qué técnicas ayudan a un aumento de las ventas de una marca en concreto?

Afortunadamente, existe una gran variedad de herramientas de software que podemos utilizar en la analítica de datos: consultas e informes, visualización, minería de datos, streaming de audio, vídeo y fotografía, entre otras. Podemos considerar por tanto a la analítica de datos como la ciencia de investigar los datos con el objetivo de obtener conclusiones sobre la información contenida en ellos, lo que es de vital importancia para la toma de decisiones en una empresa. (Joyanes L, 2014).

La analítica de datos ya se utiliza en muchas industrias para permitir a las empresas a mejorar en la toma de esas decisiones. El término de analítica de datos se usa con mucha frecuencia en el en el campo de la inteligencia de los negocios. Así, nos podemos encontrar cinco grupos distintos en la analítica de datos. (Joyanes L, 2014):

- ✓ **Analítica web**, análisis del tráfico en un sitio web.
- ✓ **Analítica social**, análisis de los datos en los medios sociales.
- ✓ **Analítica móvil**, tiene por objetivo analizar los datos que envían, reciben o transmiten los dispositivos móviles.
- ✓ **Analítica de Big Data**, analizan grandes volúmenes de datos.
- ✓ **Analítica de datos**, analizan datos transaccionales y operacionales en organizaciones o empresas.

Para conocer la importancia de la analítica de datos vamos a ir viendo, a continuación, ejemplos de casos en diferentes sectores.

4.1. Analítica de datos en comercio minorista o retail.

Las posibilidades que nos aporta la analítica de datos a las personas que se dedican al comercio minorista pueden ser múltiples. Detallamos algunos ejemplos de ellos. (Wiboanalytics, 2017).

❖ Planificación y disposición de productos.

Para conseguir un aumento de ventas se debe crear un ambiente o una atmósfera que fomente nuestro objetivo, por eso recomendamos llevar a cabo un diseño adecuado y una correcta colocación de los artículos en la tienda. Siguiendo este consejo nos permitirá determinar el tamaño de cada estantería y calcular el área de nuestra tienda, nos permitirá calcular la ubicación y la distancia con respecto a la entrada y salida de la tienda así, como conocer que artículos llaman la atención del cliente en la entrada del local.

Gracias a la analítica de datos es posible determinar qué áreas de la tienda atraen un mayor tránsito, qué productos colocar al lado de otros o incluso cómo colocarlos de tal manera que atraiga un mayor número de clientes.

❖ Optimización del surtido.

Gracias a la analítica de datos podremos ordenar los artículos de las tiendas por categoría, a raíz de esto, podemos apreciar cómo colocar de manera más efectiva cada producto en el estante correspondiente y distribuir los diversos artículos para llamar el interés de los consumidores. Para la distribución podemos seguir patrones en función de su volumen de ventas o de las preferencias de los clientes.

❖ Previsión de ventas.

Para poder tener una buena previsión de ventas dependemos de una buena realización de la analítica de datos. Esta herramienta funciona con técnicas de estadísticas que son capaces de pronosticar la demanda de nuestros productos.

❖ Gestión del inventario.

Este proceso es uno de los más importantes para tener un buen suministro, de esta forma te aseguras no perder la venta debido a la falta de stock. Gracias a ello conseguiremos racionalizar productos, optimizar la gestión del almacén y mejorar los recursos disponibles en la empresa. Para una buena gestión del inventario, la analítica de datos tiene en cuenta los tiempos de espera por parte del proveedor, el cual asegura la disponibilidad del producto en el momento establecido.

❖ Análisis de fidelización.

Debido a la competencia que hayamos a la hora de abrir un nuevo negocio, resulta muy costoso, además de un procedimiento complejo captar diversos clientes. Debido a la analítica de datos seguimos las siguientes pautas:

- Identificar los clientes más fieles sobre una marca en concreto.
- Analizar todo movimiento de los consumidores para poder ofrecerles lo que necesitan.
- Revelar distintos comportamientos de los consumidores.
- Fraccionar cada tipo de cliente para diseñarles ofertas exclusivas dirigidas a cada tipo de cliente, así se sentirán satisfechos con el servicio ofrecido.

❖ Calcular los precios.

La analítica de datos nos ofrece información útil para incrementar las ventas en función de la resolución de precios óptimos y los descuentos asociados a cada marca o el descuento aplicado para cada promoción.

❖ Análisis de la cesta de compra.

Gracias a los productos que suben los usuarios a sus cestas de compra podemos comprender cuáles son sus tendencias, hábitos e incluso los comportamientos de los mismos. Una vez que ya hayamos analizado las cestas de los consumidores podemos detectar las siguientes interpretaciones: ofrecer productos relacionados que les pueden interesar al usuario con características similares y con un precio más reducido y podremos clasificar a los clientes según los productos añadidos a la cesta.

❖ Reporting y el análisis del negocio.

El reporting es considerado como una opción muy ventajosa que se puede extraer con la analítica de datos e interactuando con Big Data obtenemos grandes conocimientos para determinar la potencia de nuestro negocio.

Podremos obtener un resultado con poco margen de error basándose en información objetiva respecto al conocimiento adquirido. Según la analítica de datos el proceso sería el siguiente: recogida de datos, almacenamiento de dichos datos, prácticas de algoritmos estadísticos de segmentación, diseño de modelos que ayuden a defenderse de los cambios de mercados, elaboración de diversos grupos de clientes y optimización de diversas labores comerciales.

4.2. Analítica de datos y Big Data en la banca.

La inteligencia artificial o también conocida como “Machine Learning” es manejada por las empresas para poder llevar un seguimiento de los patrones y los comportamientos de un cliente. Hoy en día el “Deep Learning” es una rama dentro del Machine Learning, intenta conseguir que el aprendizaje del software sea más soberano, obviando así la intervención humana. (Bbvaopenmind, 2018).

La sinergia entre estas tres tecnologías: analítica de datos, Big Data y machine learning capacitan a las empresas a renovar todas sus estructuras. En el sector financiero, las grandes empresas que se encuentran inmersas en ello han ido apreciando la enorme cantidad de datos que genera cada uno de sus clientes a lo largo del día y está llevando acciones para extraer ese conocimiento y convertirlo en información útil.

A continuación, mostramos un listado de acciones que permiten adquirir ventajas competitivas al sector financiero y para así, poder diferenciarse de los competidores: (Bbvaopenmind, 2018).

A. Cuenta inteligente.

Gracias a la cuenta inteligente, la entidad financiera se propone como objetivo ofrecerle al cliente un concepto distinto de la cuenta, pero con un atractivo adicional gracias al cual puede predecir gastos futuros, analizar los comportamientos según los gastos generados, comparar gastos con un cliente que tenga un perfil similar o que sea exactamente el mismo e incluso le puede recomendar los productos que pueden llevar en cubrir las distintas necesidades del cliente.

B. Productos financieros personalizados.

Cada usuario lleva a cabo sus propios movimientos económicos y gracias a la analítica de datos podemos descubrir distintas conductas para ofrecer productos financieros personalizados a los clientes.

C. Nuevas oportunidades de negocio sobre clientes de la entidad.

La entidad financiera cuenta con los movimientos económicos de cada cliente y también dispone de información externa del mismo, por ejemplo, si un cliente coloca en una red social unas fotos de un vehículo y el cliente lo desea, la entidad financiera puede crear una oferta para ese producto.

D. Nuevas oportunidades de negocio para personas que no son clientes.

Una entidad es capaz de detectar necesidades financieras para cada tipo de usuario, debido a esta característica permite captar nuevos clientes para la organización.

E. Gestión de riesgos y prevención de fraude.

Estos aspectos son los más importantes para cualquier entidad financiera, poco a poco ha ido mejorando gracias a los usos de analítica de datos, Machine Learning y Big Data.

F. Recomendador interno de dónde ubicar físicamente la sucursal.

Este aspecto también es importante debido a que nos ayuda a elegir la ubicación de la sucursal, para ello se tiene en cuenta varias cosas, por ejemplo, en qué zona de la ciudad ubicarlo, qué tipo de clientes son o cuántos bancos hay ubicados en una respectiva zona.

G. Recomendador interno de dónde ubicar un cajero automático.

Enlazando con el apartado anterior la entidad debe analizar las zonas en donde los clientes gasten más, ya que estos cajeros ofrecen la comodidad de sacar dinero sin hacer esperas y lo más rápido posible.

H. Analizar cuánto dinero ingresar en el cajero en caso de fines de semana y festivos.

Esta característica te permite destacar de tus competidores, si calculamos la cuantía óptima para ingresar en los cajeros, seguiremos cumpliendo el objetivo de no interrumpir nuestro servicio.

I. Predecir cuándo un cliente abandonará la entidad.

Gracias a la analítica de datos, una entidad puede predecir cuándo un cliente abandonará la entidad, este proceso se puede analizar de diversas formas, observando movimientos en su cuenta, examinar si ha seguido a otras entidades financieras mediante sus redes sociales o si hace tiempo que ha visitado la página web. Ante esta situación, la entidad puede actuar ofreciéndole al cliente diversos productos u ofertas que le ayuden a no abandonar la entidad.

4.3. Analítica web y social.

La analítica web se considera una disciplina de la analítica de datos cuyo centro de estudio son los datos que emanan a través de páginas web. La definición de analítica web según Kaushik, (2011): “Es el análisis de datos cuantitativos y cualitativos de su sitio web y de la competencia, para impulsar una continua mejora de la experiencia online que tienen tanto los clientes habituales como los potenciales y que se traduce en unos resultados esperados”. El análisis se sujeta con una serie de secuencias de clics que realizan los usuarios. Gracias a estos clics podemos conocer casi todo a cerca de los mismos y, por tanto, también permite saber qué acciones realizar para mejorar. La empresa debe tomar la decisión de si el análisis web se realiza por la misma empresa o se contrata a un proveedor externo,

Debido a la complejidad de la elección, la empresa debe examinar diferentes opciones, costes, calidad en los resultados e incluso el tiempo que disponen para implementarlo. Kaushik, (2011), confirmó que las herramientas Google Analytics, Yahoo! y Web Analytics, eran soluciones eficientes y, además, son gratuitas. Por tanto, no se debe pagar por la analítica web cuando sus necesidades no son demasiado complicadas, aunque se deberá destinar personal para la realización de estos análisis. Todas las herramientas mencionadas cumplen la función de analizar la web, sin embargo, la más recomendada y la más utilizada es la de Google Analytics.

“Podemos definir a la analítica social como la especialidad que apoya a las empresas a investigar, a calcular y a explicar el rendimiento de las iniciativas de social media en el contexto de objetivos empresariales específicos”, (Lovett, 2012). Consideramos a la analítica social como el proceso donde se mide, analiza o interpreta los datos sociales que se extraen de los diferentes canales, medios de comunicación y dispositivos.

Este análisis nos permite analizar métricas para calcular con exactitud éxitos, fracasos o situaciones de peligro para el negocio e interacciones con los clientes, todo esto nos ayuda a entender cómo el usuario percibe nuestra marca o como responden a lanzamientos de nuevos productos y campañas de marketing. El análisis social nos permite usar los datos para conocer cómo están funcionando los diferentes modelos de negocio que lleva la compañía y cómo podemos mejorarlo, es decir, la analítica social también nos aporta datos que transformamos en información para tomar decisiones correctas. (Joyanes L, 2014).

Podemos considerar las etapas de la analítica web en las siguientes etapas: captura de datos, realización del análisis, información de los resultados y acciones que puedan beneficiar a la empresa, a sus clientes y a sus socios. (Joyanes L, 2014).

A modo de ejemplo vamos a señalar dos casos en los que se utilizó Twitter para recopilar datos. En septiembre de 2012, el BBVA lanzó una página web para poder analizar en Twitter la opinión de los valores del IBEX (índice de referencia en la bolsa española). La página web se llamaba Stockbuzz.es. Dicha página captaba las opiniones de los usuarios que emitían sobre el IBEX 35 en Twitter, con el objetivo de conocer las expectativas sobre el mercado a raíz de la información que se publicaba en la red social. Mediante algoritmos automáticos se le asignaban un valor en función de su relevancia y después se agrupaban según fueran positivos, negativos o neutros. (Joyanes L, 2014).

El Grupo de Investigación de Procesamiento del Lenguaje Natural y Sistema de Información (GPLSI) de la Universidad de Alicante, en España, presentaron la herramienta TwitsObserver. Esta herramienta permitía valorar las opiniones de los usuarios de Twitter sobre un tema determinado de forma automática. (Joyanes L, 2014).

- Indicadores clave del rendimiento y métricas.

Un indicador clave del rendimiento, también conocido como KPI, es una métrica que ayuda a medir el grado de adquisición de los objetivos fijados para un sitio web. Una métrica es, una medida cuantitativa que nos permite conocer el estado de una página web, (Kaushik, 2011).

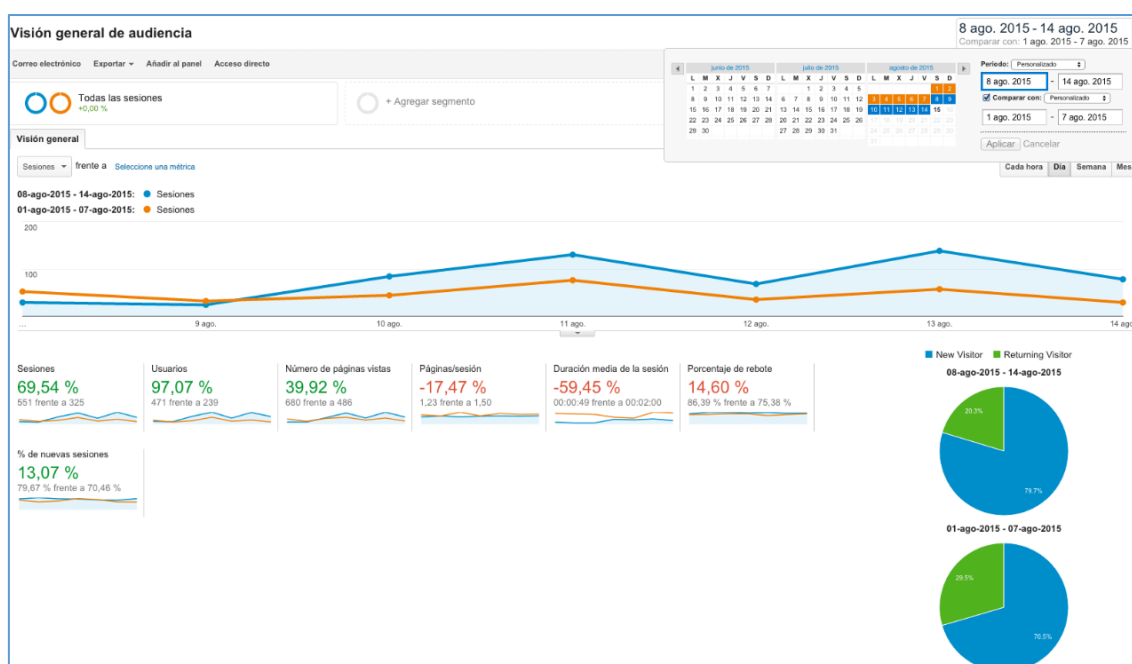
Una vez que hayamos establecido la estrategia de la empresa se deberá deducir cuáles son las métricas que deberán convertirse en KPI, dicho con otras palabras, se deben seleccionar cuáles son las métricas que ayudan a alcanzar los objetivos del negocio. Podemos mencionar algunos ejemplos de indicadores KPI. (Joyanes L. 2014):

- Para el caso de un supermercado: valores totales por venta cada hora, ventas por metro cuadrado, importe medio de factura.
- Para un comercio electrónico serían: tasa de abandono, tiempo de permanencia en la página, horario y día de la semana con mayores visitas
- Página corporativa de una empresa: número de formularios enviados, índice de incidencias atendidas o índice de descargas de documentos.

- Blog: número de comentarios, número de veces compartido, (si es alto quiere decir que el blog funciona bien, si por el contrario, es bajo, puede suponer que el funcionamiento del blog no es el adecuado).
- Reservas de viaje: tasa de conversión, tasa de rebote y tasa de permanencia. Una tasa de rebote baja y un elevado tiempo de permanencia indica qué productos son más demandados o cuáles tienen una mejor oferta.

Es importante conocer dos conceptos importantes en el análisis web: los informes y la segmentación. Los informes son datos que proporciona la herramienta de la analítica y gracias a ellos obtenemos un análisis significativo para tomar decisiones correctas. El ejemplo más claro es el de Google Analytics donde nos proporciona una cantidad increíble de informes de todo tipo y bastante extensos donde disponemos de datos fiables de casi todos los indicadores en el análisis del tráfico del sitio web. (Joyanes L, 2014). A continuación, mostraremos un ejemplo de un informe, en el cual, podemos observar el rendimiento de un blog.

Figura 4: Ejemplo de informe de Google Analytics.



Fuente: Blogpocket, 2014.

Como podemos observar en la figura 4, nos damos cuenta de que el contenido del blog no les parece atractivo a los usuarios, sí que es cierto, que ha habido un pequeño crecimiento de nuevos visitantes, aun así, no conseguimos conservarlos.

La segmentación consiste en aplicar filtros para analizar una parte específica de los datos. Desde el punto de vista de la analítica, la segmentación de visitas y visitantes es una de las técnicas que más se usan en el análisis web de una empresa. Debido a la segmentación conocemos mejor el comportamiento de ciertos tipos de usuarios, y gracias a la segmentación avanzada, introducida por Google Analytics, consiguieron un gran avance en el campo de la analítica. (Joyanes L, 2014).

Gracias al programa de Google podemos segmentar por filtros de perfil o con segmentos avanzados. La principal diferencia es, que los filtros nos va a ofrecer la posibilidad de alterar los datos existentes, en cambio, en los segmentos, esta opción no estaría disponible. Ahora vamos a mostrar unos ejemplos de segmentación, gracias al programa de Google Analytics.

Figura 5: Ejemplo de un segmento avanzado de Google Analytics.

The image shows a screenshot of the Google Analytics 'Visión general de público' (Audience Overview) interface. At the top, there is a navigation bar with the following options: 'Segmentos avanzados' (Advanced Segments), 'Correo electrónico' (Email), 'Exportar' (Export), 'Añadir al panel' (Add to panel), and 'Acceso directo' (Shortcut). Below this, there is a text input field labeled 'Nombre:' (Name). The main content area displays a configuration for an advanced segment. It starts with a dropdown menu set to 'Incluir' (Include), followed by a dropdown menu set to 'Acción de evento' (Event Action), and then a dropdown menu set to 'Que contienen' (That contain). Below these, there is a text input field, followed by the text 'o bien' (or), another text input field, and a button labeled 'Añadir instrucción "OR"' (Add "OR" instruction). Below this, there is a text input field, followed by the text 'y' (and), another text input field, and a button labeled 'Añadir instrucción "AND"' (Add "AND" instruction).

Fuente: Websa100, 2018.

Figura 6: Ejemplo de un filtro de perfil de Google Analytics.

The screenshot shows the 'Perfiles' (Profiles) section of the Google Analytics interface. It features a navigation bar with 'Perfiles', 'Información de seguimiento', 'Configuración de la propiedad', and 'Configuración social'. The main heading is 'Nuevo perfil de informes'. Below this, there is explanatory text about creating a profile and a question: '¿De qué datos debe realizar el seguimiento este perfil?'. Two options are presented: 'Sitio web' (selected) and 'Aplicación'. The 'Sitio web' option includes the text: 'Realizar el seguimiento de los sitios web cuyo código HTML controla'. The 'Aplicación' option includes: 'Realizar el seguimiento de las interacciones en las aplicaciones Android e iOS'. Below this is the 'Configuración del perfil' section, which includes a text input for 'Nombre del perfil de informes' (containing 'Mi nuevo perfil de informes'), a dropdown for 'Zona horaria de informes' (set to 'Estados Unidos'), and a time zone selection (set to '(GMT-08:00) Hora del Pacifico'). At the bottom, there are two buttons: 'Crear perfil' and 'Cancelar'.

Fuente: Websa100, 2018.

En resumen, podemos afirmar que Google Analytics es una herramienta bastante completa y útil para llevar un buen análisis y lo más importante, es completamente gratuita. Kaushik no se equivocó cuando afirmó que esta herramienta era la más eficaz.

5. ANALÍTICA DE BIG DATA

5.1. Importancia Big Data.

Por todo lo mencionado con anterioridad creemos que la correcta utilización de los datos para convertirlos en información es importante en la toma de decisiones de la empresa. Esto se hace más notable cuando la continuidad de datos es inmensa. Por eso también creemos que el Big Data es beneficioso para muchas organizaciones, ya que pueden aportar respuestas a las diversas preguntas que éstas tengan o pueden tener ofreciéndoles un punto de apoyo.

La compilación de enormes cuantías de datos y la búsqueda de los mismos hacen que las empresas produzcan cambios más rápido, les faciliten el trabajo y sean a su vez más eficientes.

El análisis de Big Data concede a las organizaciones a beneficiarse de sus datos y a emplearlos para hallar nuevas oportunidades. Todo ello conlleva a descubrir negocios más inteligentes, a conseguir unos incrementos en los beneficios y a obtener clientes más fieles. Se ha comprobado que las empresas que tienen más éxito con Big Data obtienen un aumento de valor de la siguiente forma. (PowerData, consultada en octubre 2018):

- **Disminución de costes.** Gracias a la herramienta de Hadoop, nos ofrece importantes ventajas para ayudarnos en reducir los costes a la hora de almacenar grandes cantidades de datos, e incluso, gracias a ellos, identifican métodos más eficientes para realizar negocios.
- **Más eficiente para tomar decisiones.** Debido a Hadoop o a la analítica en memoria junto con la ventaja de analizar diversas fuentes de datos, las empresas tienen suficiente capacidad para analizar la información rápidamente y tomar una decisión fundamentándose en lo que han analizado.
- **Nuevos productos y servicios.** Gracias a la analítica de los grandes volúmenes de datos, cada vez más empresas están diseñando nuevos productos o servicios para satisfacer a las exigencias de los usuarios.

Pondremos ahora algunos ejemplos de cómo puede ayudar el Big Data en diferentes sectores. (PowerData, consultada en octubre 2018):

- I. **Turismo.** Uno de los aspectos más complejos de este sector es mantener a los clientes satisfechos, aunque esta medida sea difícil de cuantificar. Sin embargo, gracias a los distintos análisis de Big Data, ofrecen a este sector la ventaja de recolectar datos de los usuarios, aplicar estos análisis y detallar distintos inconvenientes antes de que sea irremediable.
- II. **Empresas manufactureras.** Estos tipos de empresas colocan sensores en sus artículos para poder obtener datos mediante telemetría. Esto se utiliza para ofrecer ayuda a las comunicaciones, seguridad o navegación. Gracias a ello, se pueden revelar patrones de uso, porcentajes de fracaso y otras ocasiones de mejora de productos capaces de disminuir los costes de desarrollo o montaje.
- III. **Publicidad.** La mayoría de los dispositivos avanzados y otros instrumentos GPS ofrecen a las organizaciones la posibilidad de dirigirse a los consumidores y ofrecerles lo que realmente quieren, basados en su historial de búsqueda. Esto conlleva a la posibilidad de aumentar los ingresos de la empresa.
- IV. **Medicina.** Éste es uno de los sectores más importantes, ya que está consiguiendo reducir los elevados costes en la investigación clínica y ayudando a los profesionales en la toma de decisiones.

“En el ámbito de la farmacia hospitalaria, la tecnología Big Data ayudaría a comprender mejor la utilización de los medicamentos y los integraría en el conjunto del hospital para un conocimiento mayor en el uso de recursos”. Otro ejemplo de este sector sería el caso de las pandemias, ya que gracias a este análisis se puede descubrir el riesgo de una pandemia en tiempo real. (Monleón T, 2017).

5.2. ¿Qué valor tiene para la empresa el Big Data?

Anteriormente, la empresa invertía elevadas cantidades de dinero para analizar grandes cantidades de datos y, además, seguían un proceso que resultaba ser poco efectivo. Pero a día de hoy, existen soluciones para este problema y gracias a esto, se pueden convertir en un valor añadido para las mismas. Con el Big Data puede ayudar a enriquecer las decisiones de la empresa, auxiliar a las empresas en el cumplimiento de los objetivos y ayuda a gestionar los riesgos futuros. (Itsitio, 2018)

Una vez que somos conscientes de la importancia del uso de Big Data, debemos mencionar las ventajas competitivas que adquieren las empresas con él para que podamos ver en qué se diferencian de sus competidores.

La primera de ellas es la introducción de avances tecnológicos para que nos ayuden en la obtención de datos para poder revelar las distintas necesidades de la organización. La segunda ventaja es que, gracias a los análisis de los datos, las empresas tendrán mayores facilidades a la hora de tomar decisiones y podrán reducir los riesgos de manera notable. La tercera ventaja competitiva es que permite que las empresas evalúen sus productos, gracias a ello, las empresas obtienen información muy útil y les permite modificar productos ya elaborados o diseñar nuevos productos. La cuarta ventaja competitiva establece la fragmentación de los clientes para poder llevar distintas acciones para cada tipo de usuario y por último, y no menos importante, el Big Data ofrece una proximidad y facilidad para acceder a la información dentro de la empresa, de esta forma se consigue un trabajo mucho más eficiente. (Silicon, 2013).

“Cada 24 horas, los miles de millones de dispositivos y seres humanos conectados a internet generan 2,5 millones de Terabytes de información”, según IBM 2016. Para entender estos datos debemos hacer una mención a tres características principales de los datos en el Big Data que son: la velocidad, el volumen y la variedad. Con el paso del tiempo el Big Data fue desarrollando otras dos características, que son: veracidad y valor. (Joyanes, L. 2014).

Vamos a estudiar cada una con más detalle, (Joyanes, L. 2014):

- **Volumen:** cuando nos referimos al volumen, hablamos de la gran cantidad de datos que son manejados por las empresas. Se espera que, en 2020 se alcance los treinta y cinco zettabytes (ZB) en el mundo, (Joyanes. L, 2014). Sin embargo, sólo una pequeña cantidad de datos son analizados, a modo de ejemplo tenemos Twitter, que genera más de nueve terabytes (TB) cada día.

Las empresas tienen que hacer frente a una inmensa cantidad de datos, aun así, existe tecnología que nos permite analizar todos esos datos con el objetivo de un mejor entendimiento de nuestros negocios.

- **Variedad:** con el término de variedad, nos referimos a que los datos usados para el Big Data provienen de diversas fuentes, aunque no suelen llevar una estructura tradicional, por ejemplo, los datos de las redes sociales pueden proceder de una fuente de sensores y no están diseñadas para incorporarlas en una aplicación. Existen diversos riesgos por no invertir en Big Data, por ejemplo: disponer de una gran cantidad de información puede imposibilitar ver oportunidades dentro de nuestro negocio, o incluso, impide ver las amenazas dentro del mismo.
- **Velocidad:** esta característica se refiere a la rapidez con la que se generan los datos. Por consiguiente, el procesamiento y el posterior análisis de los datos debe hacerse en tiempo real.
- **Veracidad:** esta característica señala la obligación que tienen las organizaciones de que los datos obtenidos sean relevantes y les ayuden de forma eficiente a la toma de decisiones. La veracidad supone un desafío para el Big Data debido al rápido crecimiento de las fuentes de datos.
- **Valor:** las organizaciones persiguen el objetivo de obtener la información de manera eficaz, para ello, utilizamos las tecnologías de código abierto, como, por ejemplo, Hadoop. De esta forma, se obtendrá mejor información y nos ayudará de manera eficiente en la toma de decisiones, gracias a ello, conseguiremos incrementar nuestro valor como empresa.

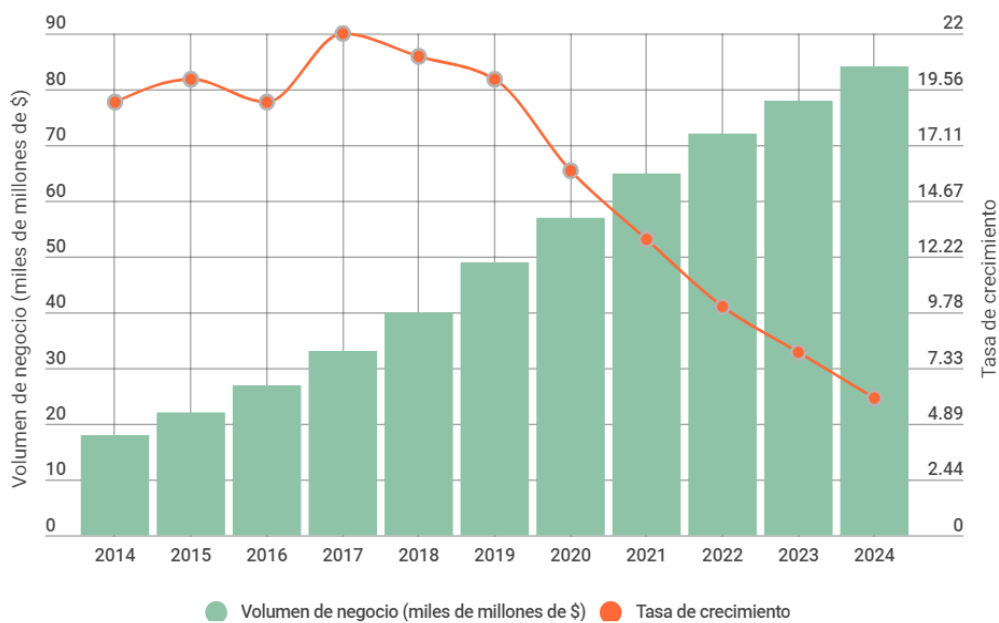
“Las compañías estadounidenses perdieron 3,1 millones de dólares por culpa de una mala gestión de los datos, según informa IBM”, 2016. Tenemos que añadir, que según Harvard Business Review, (2016) “los llamados trabajadores del conocimiento emplean un 50% de su tiempo de trabajo buscando datos o corrigiendo errores por no confiar en la veracidad de los datos suministrados”. Desde Forbes Corporation Communications, (2017), generaron un resumen de aspectos negativos que pueden ocasionarse debido a una mala gestión de los datos, destacando:

- A. Impacto en la reputación.** Dependiendo del negocio, una mala gestión de los datos puede generar una mala imagen en la empresa que puede afectar al interés que muestran los clientes y los consumidores.
- B. Oportunidades desaprovechadas.** Un sesgo en el seguimiento del comportamiento del consumidor puede dejar posicionada en mal lugar a la empresa, lo que conlleva una desventaja frente a la competencia.
- C. Decrecimiento de ingresos.** Si realizamos una mala gestión de los datos, iremos arrastrando operaciones comerciales que nos pueden ocasionar pérdidas de gran relevancia.

“Una parte de proyectos de Big Data fracasan en España al iniciarse sin un objetivo claro. Algunas empresas realizan estos proyectos únicamente por subirse a la “ola” de lo que es la tendencia, pero desconocen qué quieren hacer exactamente, ni cómo lo van a hacer y ni qué necesitan para lograrlo”, según Raúl Galán en 2016, CTO de Hocelot, una consultoría especializada en Big Data.

A pesar de los riesgos que tenemos que tener en cuenta, este tipo de negocio ofrece un interesante volumen de beneficio. En la figura 1 se observa que este mercado aumentará más del 10% anual hasta 2022, posteriormente se ralentizará, como cualquier mercado que se encuentra en su periodo de madurez, aunque no se detendrá. Entre 2014 y 2024, se estima que habrá multiplicado su tamaño por seis.

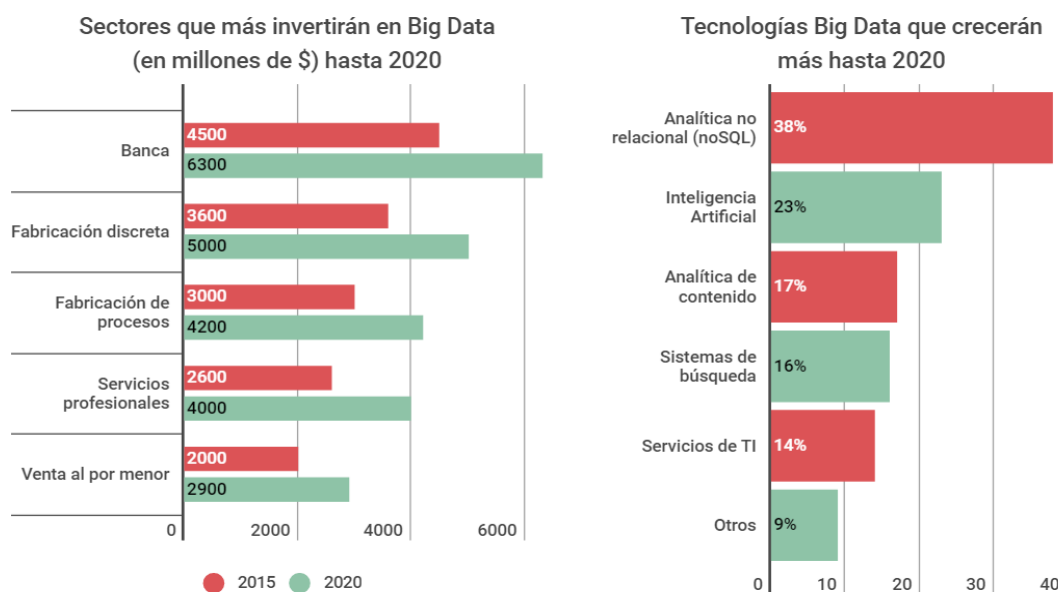
Figura 7: El mercado de Big Data en el mundo.



Fuente: hablemosdeempresas.com (2018).

Asimismo, algunos sectores obtendrán un mayor beneficio por el análisis de datos en los años venideros, por ejemplo, la banca o la producción en fábricas, sin embargo, no serán los únicos beneficiados, como mostramos en la siguiente figura.

Figura 8: Sectores y tecnologías que impulsarán la inversión.



Fuente: hablemosdeempresas.com (2018).

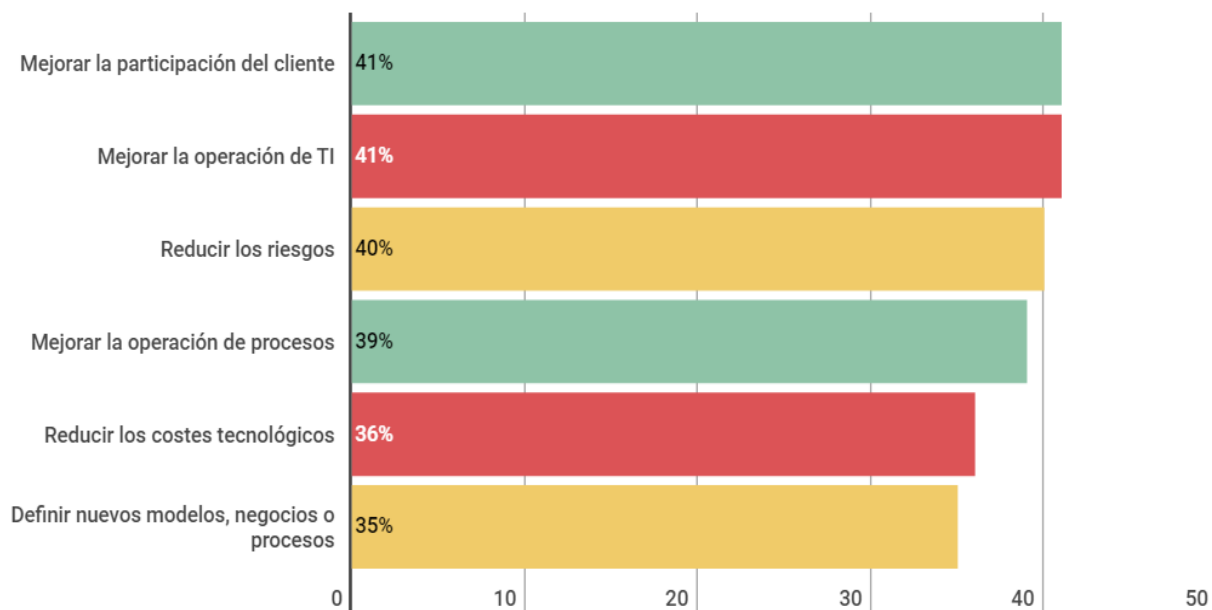
Según Carmen Reina (2016), de Orange España, “el sector de las telecomunicaciones es uno de los más implicados, al ser uno de los que más volumen tienen de datos. Ahora mismo, las telecomunicaciones están muy centradas en satisfacer al cliente, en retenerlo. Nuestro objetivo ya no es dar un producto, sino que el cliente esté satisfecho.”

En un estudio realizado por CaixaBank (2017), “en el sector financiero, muchos bancos e incluso aseguradoras tienen iniciativas de Big Data en marcha, permiten incrementar el conocimiento de los clientes para mejorar la oferta comercial, personalizar el servicio teniendo en cuenta las preferencias de los clientes, incorporar nuevos datos en los modelos de riesgo y simplificar procesos”.

En España, BBVA, como ya hemos mencionado, es uno de los bancos que con más decisión ha apostado fuertemente por la tecnología. Esta entidad financiera cuenta con la plataforma RedeX, cuya misión consiste en afinar la evaluación de riesgos cuando un cliente nuevo o existente pide un crédito bancario. Esta plataforma tan compleja usa diversos datos de transferencias y otras operaciones con terceros para que los expertos puedan visualizar el impago o la morosidad.

Arturo Larraínzar (2017) de la empresa Atresmedia señala, “que una empresa explote bien o mal el Big Data depende mucho de sus capacidades y su cultura a la hora de analizar. Creo que lo que hay hoy en día son muchas empresas que estamos tratando de ver cuáles son las posibles explotaciones del Big Data y aprendiendo”.

Figura 9: ¿Qué quieren conseguir las empresas españolas con el Big Data?



Fuente: hablemosdeempresas.com (2018).

Como hemos observado en la figura 6, las empresas españolas buscan con el Big Data en mejorar las tecnologías de la información, es decir, almacenar y manipular datos para adoptar mejores decisiones.

Por esa razón, las organizaciones procuran reducir riesgos gracias al Big Data. No obstante, las empresas persiguen el objetivo de definir nuevos modelos de negocio, pero en menor medida que las que hemos mencionado anteriormente.

5.3. ¿Cómo crear valor para los clientes a través de Big Data?

El Big Data permite a las empresas llevar una experiencia más enriquecedora en el trato de sus clientes. Los beneficios que aportan a sus usuarios son múltiples, por ejemplo: ofertas adaptadas a los intereses de los mismos o una atención más personalizada, gracias a ello, los usuarios se mostrarán satisfechos con la ayuda recibida. El sector que más valor aporta Big Data para los usuarios es el de la medicina, ya que, con un buen sistema de gestión de datos puede ayudar a los profesionales a atendernos de manera más eficiente. (Hablemosdeempresas, 2018).

Continuamente los esfuerzos de los departamentos de marketing son cada vez más grandes. Se podría decir que su objetivo se centra en ilustrar un retrato cada vez más detallado del consumidor, registrando todo tipo de referencias en medios de comunicación, registrando sus hábitos de compra y analizando lo que introduce en la cesta de compra en una página de compras online como puede ser por ejemplo “Aliexpress”. Otras analizan intereses o aspiraciones. Gracias a todo ello, las empresas consiguen un retrato robot del perfil del consumidor y se encargan de ofrecerles productos que realmente necesitan o que encajarían con sus perfiles de consumidores. (Saimasolutions, 2016).

Muchas de las grandes empresas actuales como, por ejemplo: TripAdvisor, Uber, EBay, Netflix o Amazon, consiguen recopilar enormes cantidades de datos de sus clientes; ya sean cuentas personales o datos bancarios, datos sobre las pulsaciones en el teclado o datos de localización e incluso valoraciones del servicio tanto de proveedores como de los propios consumidores. Como todos nosotros, a la hora de comprar un producto lo primero que hacemos es buscarlo en distintas páginas y a continuación leemos las valoraciones, y es aquí donde mostramos más o menos interés por el producto debido a que esas valoraciones son importantes para la adquisición del mismo y ayudan a la decisión de compra por parte del usuario. Gracias a Big Data podemos analizar todos esos datos y convertirlos en ventajas competitivas para poder ofrecerle al cliente la mayor seguridad al comprar por páginas web. (Saimasolutions, 2016).

5.4. ¿Cómo afecta Big Data a la sociedad en general?

“No solo son importantes los datos y el conocimiento que nos aportan los mismos, sino que están cambiando la economía mundial” Monleón, (2010). En 2014 la Comisión Europea mostró una novedosa estrategia basada en Big Data para dar soporte y así acelerar el cambio para llegar a una economía fundamentada en los datos. Gracias a este proceso se conseguiría la estimulación de la investigación y de la innovación en general, y no sólo eso, sino que también abrirá las puertas a nuevas oportunidades de negocio y la posibilidad de que las pequeñas y medianas empresas (PYME) dispongan de un aumento de capitales para ayudarles a ir creciendo.

Dichas declaraciones sobre la Unión Europea se recopilan en el artículo “Towards a thriving data-driven economy”, (2014), donde se indica que esperan que la tecnología y los servicios ofrecidos por el Big Data crecerán en todo el mundo a un crecimiento anual del 40% aproximadamente, unas siete veces a las que crece el mercado de las Tecnologías de la Información y Comunicación (TIC).

¿Cómo puede mejorar nuestra vida ante esta recopilación tan masiva de datos? Como dice Sánchez (2013), “las grandes compañías tecnológicas disponen de centros de almacenamiento para guardar estas grandes fuentes de información y tras su análisis pueden estudiar el comportamiento de los clientes para realizar acciones comerciales más efectivas o focalizar la publicidad a los intereses del consumidor”.

La Comisión Europea publicó un interesante escrito en 2014 titulado: “Making big data work for Europe” en el cual explica con detalle la importancia del Big Data. Hoy en día, los datos son un instrumento importante para la economía y nuestra sociedad. Gracias al Big Data se están dando lugar a innovaciones tecnológicas, se desarrollan nuevas herramientas y también se desarrollan nuevas habilidades. Si las empresas son capaces de proporcionarles un buen uso de estos datos pueden llevar consigo oportunidades para sectores tales como: el transporte, la salud, la banca, las telecomunicaciones o la fabricación. La Comisión Europea mostró algunos ejemplos de cómo el Big Data cambiaría la sociedad:

- A. Transformación de las industrias de los servicios en Europa, gracias a la creación o modificación de productos o servicios renovadores.
- B. Incremento de la productividad de los diversos sectores.
- C. Enriquecerán de manera eficiente tanto las investigaciones como las innovaciones.
- D. Decrecimiento de costes significativas.

Otro aspecto en el que proporciona ayuda el Big Data es al económico. Un ejemplo de este tema es el Massachusetts Institute of Technology (MIT) recoge datos sobre precios de los productos que son vendidos en internet y gracias a ellos estima tasas de inflación. (Monleón T, 2017).

5.5. La importancia del Big Data en los negocios.

Como hemos comentado con anterioridad, el papel del Big Data en los negocios es incuestionable. Sin embargo, los directivos pueden no tener los datos correctos o pueden que no sean capaces de interpretarlos o también puede ser que no sean capaces de obtener los datos que realmente necesitan.

Para lidiar con este tipo de inconvenientes nace el concepto de “inteligencia del negocio” que ha ido evolucionando e integrando novedosas tecnologías y herramientas que les sirven de gran ayuda a los directivos a tomar decisiones empresariales. (Joyanes L, 2014).

“La inteligencia de negocios es una colección de tecnologías y sistemas de información que dan soporte a la toma de decisiones empresariales o el control operacional, proporcionando información de operaciones internas y externas” (Laudon, 2012). Desde la práctica, la inteligencia de negocios está compuesta por un conjunto de aplicaciones que recomiendan cómo analizar los datos del usuario, de cómo deben presentarse los resultados en sus análisis y de cómo, una vez llevados a cabo por los analistas, los gerentes y los ejecutivos implementan estos resultados.

Debido a la complejidad de las implementaciones de la inteligencia de negocios, la mayoría de los proveedores ofrecen aplicaciones con programas clásicos de gestión de empresas tales como: CRM, ERP y SCM que a su vez transitan por la web.

Para llevar a cabo una buena gestión debemos estudiar los modos en los que las organizaciones utilizan las aplicaciones de inteligencia de negocios. Los datos se almacenan en un data warehouse o data mart, después de analizar dichos datos, presentamos los resultados de los análisis y acto seguido, se deben proporcionar herramientas para implementar los resultados. Un data warehouse está considerado como una especie de almacén electrónico en el que generalmente la empresa guarda una gran cantidad de información. (PowerData, consultada en marzo 2019).

Un data mart también es un almacén, pero se centra en una única área, como por ejemplo, marketing o contabilidad, y ésta es la principal diferencia entre data warehouse y data mart. (Blog.mdcloud, consultada en marzo 2019).

Como ya hemos mencionado conviene diferenciar entre análisis y analítica. Análisis es el término general que se refiere a un proceso, mientras que la analítica se refiere a un método que utilizan los datos para aprender algo.

Gracias a inteligencia de los negocios, los análisis de datos se pueden dividir en tres grandes categorías: análisis multidimensional (OLAP), minería de datos y sistemas de apoyo a la decisión (DSS). (Joyanes L. 2014).

A. Sistemas de apoyo a la decisión.

Eom y Kim, en 2006, citaron una descripción de DSS que incorpora la reseña de distintas personas que los han investigado a través de los años: “Un DSS es definido como un sistema interactivo basado en computadores que da soporte a los tomadores de decisiones en vez de reemplazarlos, utiliza datos y modelos para resolver problemas con diferentes grados de estructura: no estructurados (Bonczek et ál., 1981), semiestructurados (Bennett, 1983, Keen and Scott Morton, 1978), tareas estructuradas y no estructuradas (Sprague and Carlson, 1982), semiestructuradas y no estructuradas (Thierauf, 1982) y se enfoca en la efectividad más que en la eficiencia de los procesos de decisión”. (Joyanes L, 2014).

Un sistema DSS se diferencia de los demás, porque lo componen los siguientes subsistemas. (Scielo, 2009).

- ❖ **Administrador de datos:** lleva consigo una base de datos con información de gran importancia para cualquier circunstancia llevado a cabo por el DBMS (sistema encargado de manejar la base de datos).
- ❖ **Administrador del modelo:** consisten en un software que incluye patrones cuantitativos, matemáticos, financieros o científicos que aporta al sistema la capacidad de analizar.
- ❖ **Interfaz del usuario:** lleva integrada una interfaz muy intuitiva.
- ❖ **Administrador del conocimiento:** actúa como componente para razonar la toma de decisiones.

Los DSS tienen la ventaja que se puede aplicar en la mayor parte de las industrias y llevan consigo beneficios para la empresa. A modo de ejemplo:

- Gracias a los DSS, permite a los administradores de las universidades programar distintos horarios de manera eficiente.
- Un DSS permite desarrollar programas sobre pronósticos de ventas.

Nos gustaría citar algunos ejemplos de actuación real de la DSS que pueden hacernos ver la importancia de este sistema:

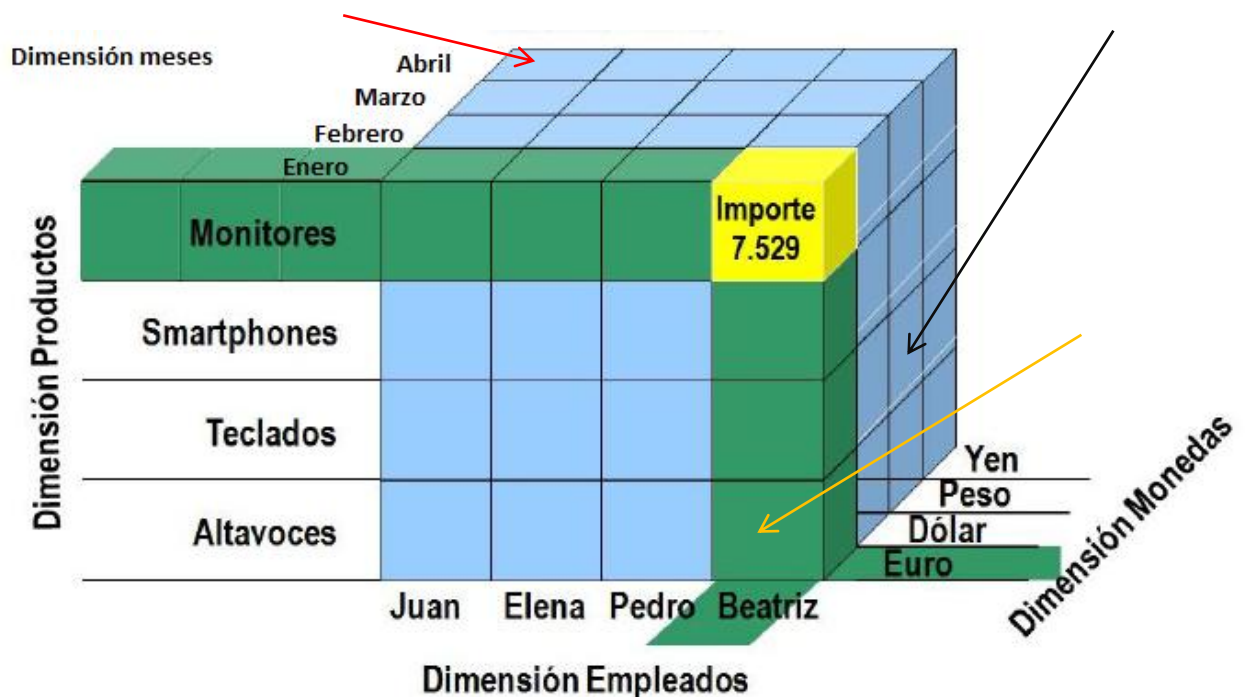
- El ejército estadounidense desarrolló un DSS del potencial del personal, cuyo objetivo era ayudar con las decisiones de reclutamiento, entrenamiento, educación y ascensos. Además, incluía características de “qué sucedería si” y pudo interactuar con una base de datos en línea y con otros programas de análisis estadístico. (Scielo, 2009).

- Cinergy Corporation, una compañía de servicios públicos de electricidad, desarrolló un DSS para reducir los tiempos de espera y los esfuerzos necesarios para el proceso de toma de decisiones en el área de la compra en el carbón. (Scielo, 2009).

B. Análisis multidimensional, cubo OLAP.

El procesamiento analítico en línea (OLAP) faculta a los usuarios a la visualización de los datos de diferentes dimensiones representados mediante un cubo multidimensional, en donde cada región del mismo representa un cierto tipo de información. A continuación, mostraremos un ejemplo de este sistema.

Figura 10: cubo OLAP



Fuente: “Big Data. Análisis de grandes volúmenes de datos en organizaciones” (2014).

En la figura 4 se muestran un cubo de datos donde se aprecia una determinada dimensión a la que hemos llamado “dimensión de empleado”. Si giramos el cubo 90 grados a la izquierda nos encontramos con otra dimensión que nos informa sobre los productos que están siendo analizados según las ventas. Si miramos desde arriba observamos los meses analizados, pero si volvemos a girar el cubo a su posición original, y esta vez, 90 grados a la derecha observaremos las diferentes monedas de una empresa internacional.

Con la flecha **negra** mencionamos al importe en pesos en el mes de marzo de los teclados y a cargo de Elena.

Con la flecha **roja** relacionamos el importe en yenes del mes de abril de los monitores y a cargo de Juan.

Con la flecha **naranja** hacemos referencia al importe en euros del mes de enero de los altavoces y a cargo de Beatriz.

Las ventajas que más destacan de los cubos OLAP son: facilitar el sistema y utilizarlo para realizar informes que ayuden a progresar en las distintas operaciones de la empresa e incluso tomar mejores decisiones, y ofrecer un método que nos permite reducir el tiempo de espera para obtener información que necesitamos. (Kyocera, 2017).

C. Minería de datos.

Definimos la minería de datos como un proceso que usa técnicas estadísticas, matemáticas e inteligencia artificial, para reconocer información útil y convertirla en conocimiento. Esta información introduce correlaciones, tendencias o modelos de predicción. (Joyanes, L. 2014).

A causa del rápido crecimiento de la información, se han convertido en una de las principales metas de la empresa, ya que, los directivos requieren una percepción más minuciosa de sus negocios. De esta forma, permite a las empresas a identificar nuevos clientes o reconocer situaciones de fraude.

Las etapas más importantes de la minería de datos son, (Morales, D. 2016):

- **Determinar los objetivos:** la minería de datos nos ofrece un amplio abanico de posibilidades, sin embargo, debemos fijar los objetivos que deseamos alcanzar.
- **Procesamiento de los datos:** esta etapa es la más importante, ya que es la que más tiempo y esfuerzo vamos a invertir. En este periodo se irán asociando los datos, se eliminan los datos que no se consideren apropiados y, además, se seleccionarán los más importantes.
- **Determinación del modelo:** como hemos mencionado anteriormente, se emplearán técnicas estadísticas y matemáticas con los que manejamos los datos, y en función, de los objetivos fijados se emplean diversos algoritmos.
- **Análisis de resultados:** comprobamos que los resultados obtenidos nos ayudarán a alcanzar los objetivos que previamente establecimos.

La minería de datos está presente en diversas áreas de negocio. A modo de ejemplo destacamos, (Morales, D. 2016):

- En el sector bancario se consiguió: predecir el pago del préstamo y analizar las posibilidades de devolución de un crédito por parte de los clientes, ofertas personalizadas para cada cliente según las características o necesidades de cada uno, detectar fraudes de blanqueamiento de dinero.
- El sector de las telecomunicaciones juega un papel importante, ya que conceden diversos estudios como: análisis de datos de las telecomunicaciones, análisis de patrones fraudulentos e incluso, reconocimiento de hábitos y tendencias.

- En el ámbito del deporte, la minería de datos se usa para prevenir lesiones, el equipo AC de Milán integra este sistema a día de hoy. Este sistema clasifica a los jugadores según su rendimiento, alimentación, forma física y entre otros. Estos datos se analizan nuevamente cada quince días.

6.CONCLUSIONES

El progreso de la tecnología ha facultado que los distintos dispositivos como; ordenadores, móviles, sensores o cámaras, estén presente en nuestras vidas. Actualmente, los datos tienen mucho protagonismo para el desarrollo de los negocios y para las empresas, su uso es cada vez más importante y es clave para la evolución de las mismas. El Big Data crea valor de diversas maneras, siempre y cuando se use correctamente y sea fácil de comprender, si los datos que se recogen son transparentes beneficiaría a muchos sectores, por ejemplo, el de la banca. Gracias a los datos, ayudan a las empresas a entender a sus clientes, pueden escuchar sus necesidades y pueden proponerles ofertas que resulten más atractivas para asegurarse la comodidad del usuario. Sabemos que el Big Data tiene un impacto en los negocios, los clientes, la sociedad y también es bastante útil para cooperar en el proceso de la toma de decisiones.

Hemos podido observar en este trabajo la importancia que tiene el Big Data en los negocios, debido a su versatilidad podemos aplicarlo a diversos sectores como la banca, la medicina o el turismo. Gracias al Big Data hemos podido contemplar las ventajas que le ofrecen, por ejemplo: productos financieros personalizados, oportunidades de negocios sobre los clientes, control del nivel de riesgo, prevenir fraudes e incluso predecir cuándo un cliente abandonará la empresa.

El Big Data está al alcance de todas las empresas. En este trabajo hemos visto los distintos programas informáticos de código abierto y gracias a la versatilidad de los mismos se pueden aplicar en diversos sectores como, por ejemplo, en Investigación y Desarrollo (I + D), en marketing, en ventas, en las telecomunicaciones, transporte y turismo.

Como consecuencia, las empresas apuestan más por trabajar con Big Data y gracias a ello pueden establecer distintos puestos de trabajo a personas interesadas en dedicarse a su análisis. Precisamente por esta razón el Big Data está creciendo de manera exponencial.

Anteriormente, las actividades que exigían más recursos estaban ligadas a empresas de gran tamaño, sin embargo, el Big Data está ajustando sus presupuestos y esto conlleva a que poco a poco se vaya adaptando a todas las empresas, ya sean grandes o pequeñas. Podemos afirmar que las oportunidades que ofrece Big Data son enormes.

El futuro de Big Data vendrá marcado por el rápido crecimiento de las fuentes de datos, especialmente por la comunicación de máquinas (M2M) y por el internet de las cosas (IoT). Estos dos conceptos se están convirtiendo en herramientas muy potentes para propulsar la actividad económica y, por tanto, serán capaces de aumentar los resultados económicos.

Resulta evidente que el Big Data aporta un valor importante a las empresas, sí que es cierto que numerosas empresas lo están empezando a integrar y es debido a las ventajas competitivas que éste ofrece, tanto sus clientes como la propia empresa se ven beneficiadas si hacemos un correcto uso del mismo, ya que existen software de código abierto que nos permiten analizar los datos sin coste alguno para estar disponible para cualquier organización. Por consiguiente, podemos establecer los objetivos de la empresa, la reducción de riesgos, aumentar la participación del cliente, definir nuevos modelos de negocio, elaboración de nuevos productos e incluso modificación de productos existentes.

El Big Data también aporta valor a los clientes o usuarios, por ejemplo, con ofertas más personalizadas o con una atención al cliente más eficaz. En nuestro trabajo hemos hablado de IoT, gracias al cual, permite llevar unos de los grandes proyectos de Big Data, las ciudades inteligentes, debido a éste permite incrementar la calidad de vida de los ciudadanos como, por ejemplo, optimizando los recursos de energía o incluso de agua.

7. BIBLIOGRAFÍA

- Consultora IDC (2012), "El universo digital de datos 2012".
- Codd E. (1970): "A Relational Model of data for Large Shared Data Banks". IBM Research Laboratory, 13, 6.
- Eom S. y Kim E, (2006): "A survey of Decision Support System Applications, The Journal of the Operational Research Society". Vol. 57, No. 11, 1264.
- Frank B. (2012): "Taming the Big Data Tidal Wave. Finding Opportunities in Huge Data Streams with Advanced Analytics". New Jersey: Wiley.
- Galán R. (2016) CTO de Hocolot: "Big Data: cómo se pasa de cantidades ingentes de datos a información valiosa que ayude en la toma de decisiones".
- IBM, Big Data y Analytics Hub (2016): "Extracting business value from the 4 V of Big Data".
- ISACA: Data Analytics (2011): "A Practical Approach" White Paper.
- Joyanes L. "Big Data. Análisis de grandes volúmenes de datos en las organizaciones" (2014). 1-5, 7-10, 102, 251-254, 260-262, 268-272, 284-285.
- Kaushik A. (2011), "Analítica web 2.0". Barcelona: Gestión 2000.
- Kaul A. (2012), revista "Information Week" edición de Mexico.
- Laudon, (2012): "La inteligencia de negocios en Big Data", 5, 101
- Morales D. (2016): "Minería de datos. Aplicaciones de técnicas predictivas". 1.1, 1-7.
- Merv A. (2011), "It's going mainstream and it's your next opportunity". Teradata Magazine, 3.
- McKinsey Global Institute, (2011). "Big Data: The next frontier for innovation, competition and productivity".
- Monleón A. (2010): "El impacto de Big Data en la sociedad de la información. Significado y utilidad". Historia y comunicación social, volumen 20, 430-431
- Monleón T. et oll "Big Data. Hacia la cuarta revolución industrial" (2017). 18-19.
- Reina C. (2016), Manager Data Scientist en Orange España: "Big Data: cómo se pasa de cantidades ingentes de datos a información valiosa que ayude en la toma de decisiones".
- Rendman C. (2016), Harvard Business Review: "Bad Data Costs the U.S. \$3 Trillion Per Year".

- Rodríguez F. (2014): “Herramientas para Big Data, entorno Hadoop” 3. 55-79.
- Sánchez J. (2013): “El impacto de Big Data en la sociedad de la información. Significado y utilidad”. Historia y comunicación social, volumen 20, 431
- Serrano P. (2014): “La tecnología ‘Big data’ podría agilizar y abaratar la gestión clínica”.
- Thaddeus Burns, (2017) Organización Mundial de la Propiedad Intelectual: “Regulación de los datos generados por máquinas: Menos es más en aras del crecimiento mundial”.

8. BIBLIOGRAFÍA DE PÁGINAS WEB

Arsys.es (<https://www.arsys.es/blog/programacion/cloud-couchdb-bbdd/>). Consultada en febrero 2019.

Acens.com (<https://www.acens.com/wp-content/images/2014/02/bbdd-nosql-wp-acens.pdf>). Consultada en noviembre 2018.

Blogpocket.com (<https://www.blogpocket.com/2014/09/01/4-informes-de-google-analytics-para-principiantes/>). Consultada en marzo 2019.

Blog.mdcloud.es (<https://blog.mdcloud.es/data-mart-que-es-y-por-que-necesitas-implementarlo/>). Consulta en marzo 2019.

Bbvaopen4u.com (<https://bbvaopen4u.com/es/actualidad/neo4j-que-es-y-para-que-sirve-una-base-de-datos-orientada-grafos>). Consultada en febrero 2019.

Blogwivoanalytics.com (<https://blog.wivoanalytics.com/anal%C3%ADtica-de-datos-en-retail-ejemplos-de-uso>). Consultada en diciembre 2018.

Big Data Internacional Campus (<https://www.campusbigdata.com/big-data-blog/item/130-predicciones-de-la-tecnologia-big-data-para-el-2019>). Consultada en marzo de 2019.

Cic.es (<https://www.cic.es/iot-sus-aplicaciones/>). Consultada en marzo 2019.

Camavalencia.com (<https://ticnegocios.camaravalencia.com/servicios/tendencias/los-efectos-positivos-de-la-analitica-de-datos-en-la-empresa/>). Consultada en marzo 2019.

Genbeta.com (<https://www.genbeta.com/desarrollo/mongodb-que-es-como-funciona-y-cuando-podemos-usarlo-o-no>). Consultada en diciembre 2018.

Garantu.com (<https://grupogaratu.com/5-beneficios-del-iot-en-la-empresa/>). Consultada en marzo 2019.

Hablemosdeempresas.com (<https://hablemosdeempresas.com/grandes-empresas/ventajas-big-data/>). Consultada en septiembre 2018.

Hablemosdeempresas.com (<https://hablemosdeempresas.com/grandes-empresas/big-data-smart-data/>). Consultada en diciembre 2018.

Ionos.es (<https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/bases-de-datos-relacionales/>). Consultada en octubre 2018.

Isciii.es (<http://scielo.isciii.es/scielo.php>) Consultada en marzo 2019.

Kyocera.es (<https://smarterworkspaces.kyocera.es/blog/ventajas-cubo-olap/>). Consultada en marzo 2019.

Microsoft.com (<http://customers.microsoft.com/en-us/story/gaelic-athletic-association-croke-park-it-technology-azure-cloud-windows10>). Consultada en marzo 2019.

Movistar.es (<http://www.movistar.es/grandes-empresas/soluciones/fichas/machine-to-machine-m2m/>). Consultada en octubre 2018.

Powerdata.es (<https://www.powerdata.es/big-data>). Consultada en octubre 2018.

Powerdata.es (<https://www.powerdata.es/data-warehouse>). Consultada en marzo 2019.

PowerData.es (<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/405808/ejemplos-pr-cticos-del-uso-de-hadoop-en-big-data>). Consultada en octubre 2018.

Paradigmadigital.com (<https://www.paradigmadigital.com/dev/cassandra-la-dama-de-las-bases-de-datos-nosql/>). Consultada en enero 2019.

Paradigmadigital.com (<https://www.paradigmadigital.com/techbiz/no-solo-clave-valor-redis-te-da-alas/>). Consultada en enero 2019.

Puromarketing.com (<https://www.puromarketing.com/42/26486/datos-sociales-son-muy-utiles-para-empresas-pero-mayoria-sigue-sin-saber-como-sacarles.html>). Consultada en enero 2019.

Saimasolutions.com (<https://saimasolutions.com/como-big-data-crear-valor-para-clientes/>). Consultada en enero 2019.

Scielo.org (<http://scielo.isciii.es/scielo.php>). Consultada en noviembre 2018.

Scielo.org (http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-56092009000200015). Consultada en noviembre 2018.

Silicon.es (<https://www.silicon.es/las-cinco-ventajas-competitivas-que-aporta-el-big-data-49286>) Consultada en marzo 2019.

Tecnologiasinformacion.com (<https://www.tecnologias-informacion.com/nosql.html>). Consultada en diciembre 2018.

Tendencias21.net (https://www.tendencias21.net/telefonica/Que-es-la-comunicacion-M2M_a801.html). Consultada en octubre 2018.

Wipo.int (https://www.wipo.int/wipo_magazine/es/2017/06/article_0005.html). Consultada en noviembre 2018.

Websa100.com (<https://www.websa100.com/blog/segmentacion-con-google-analytics/>). Consultada en marzo 2019.