



FACULTAD DE CIENCIAS ECONÓMICAS Y
EMPRESARIALES

BIG DATA EN LA BANCA Y SUS IMPLICACIONES PARA EL FUTURO

Autor: Marta Cueto Vázquez
Director: José Portela González

MADRID | Abril 2019

Marta
Cueto
Vázquez

BIG DATA EN LA BANCA Y SUS IMPLICACIONES PARA EL FUTURO



Resumen

Este estudio busca entender las diversas aplicaciones de las técnicas del *Big Data* y sus consecuencias en el modelo de negocio del sistema financiero y más concretamente, en las entidades bancarias.

Big Data se ha convertido en uno de los fenómenos más importantes de la última década. En parte se debe a que la mayoría de los procesos han sufrido un proceso de digitalización, pero también debido al surgimiento de nuevos modelos de negocio (redes sociales y blogs) y, sobre todo, a la proliferación del uso de dispositivos electrónicos y de Internet, lo cual ha potenciado la generación continua de grandes cantidades de datos. Es una realidad que el uso de Internet y su aplicación a todos los sectores de actividad, ha cambiado a todos los niveles la forma en la que operamos: desde el funcionamiento de los gobiernos centrales, las fuentes de ingreso de las empresas, hasta el modus-operandi de las personas en su vida cotidiana. La tendencia actual nos sitúa en un momento en el que la cantidad de información que se genera sobrepasa a las técnicas existentes de almacenamiento de datos. En consecuencia, todos los actores tienen que ser capaces de adaptarse a las nuevas tendencias y a la revolución en los modelos de negocio si quieren asegurarse un futuro estable.

En un primer momento, se lleva a cabo un análisis sobre las principales características del término *Big Data* de manera que permitan al lector comprender su aparición, su evolución y su relevancia en el escenario macroeconómico internacional. Por otro lado, se estudia el contexto actual de la Banca y de las aplicaciones ya hechas con las técnicas de *Big Data*. Finalmente, se formulan las perspectivas de futuro y la potencial aparición de nuevas tecnologías que todavía redefinan más el modelo de negocio de la banca tradicional para así construir un banco del futuro.

Palabras clave: *Big Data*, sistema financiero, la banca, nuevas tecnologías, redefinición de los modelos de negocio, análisis de la información, escenario internacional, perspectivas de futuro.

Abstract

This study seeks to understand the diverse features of the *Big Data* techniques and their consequences in the business model of the financial system and, more specifically, in banking entities.

Big Data has become one of the most important phenomena of the last decade. This is partly due to the fact that most of the developments have undergone a process of digitalization, but also due to the emergence of new business models (social networks and blogs) and, above all, the proliferation of the use of electronic devices and the Internet, which has enhanced the continuous generation of large amounts of data. It is a reality that the use of the Internet and its application to all sectors of activity, has changed at all levels the way in which we operate: from the operation of central governments, the sources of income of companies, to the modus -operandi of people in their daily lives. The current trend places us at a time when the amount of information generated exceeds the existing data storage techniques. Consequently, all actors must adapt to new trends and the revolution in business models if they want to ensure a stable future.

At first, an analysis is carried out on the main characteristics of the term *Big Data* in a way that allows the reader to understand its emergence, its evolution and its relevance in the international macroeconomic scenario. On the other hand, the current status of banking and the applications already made with *Big Data* techniques are studied. Finally, future prospects and the potential emergence of new technologies that still redefine the traditional banking business model in order to build a bank of the future are formulated.

Keywords: *Big Data*, financial system, banking, new technologies, redefinition of business models, information analysis, international scenario, future perspectives.

TABLA DE CONTENIDO

Resumen	I
Abstract	II
Índice de figuras	IV
1. Introducción	1
1.1 Justificación de la importancia del tema.....	1
1.2 Objetivos.....	2
1.3 Metodología.....	3
2. ¿Qué es <i>Big Data</i>?	5
2.1 Definición y conceptos (3vs y 5vs).....	5
2.2 Historia y evolución.....	7
2.3 Cómo obtener los datos.....	10
2.4 Almacenamiento de la información	13
2.5 Cómo analizar-procesar los datos	15
2.6 Cómo usar los datos para crear valor	18
3. <i>Big Data</i> en Banca	21
3.1 Motivos para la utilización de <i>Big Data</i> en la Banca.....	22
3.2 Aplicaciones en distintas áreas	26
3.3 Beneficios del <i>Big Data</i>	32
3.4 Inconvenientes y peligros del <i>Big Data</i>	34
4. Banca en el futuro	37
4.1 <i>The Internet of Things</i> (IoT)	37
4.2 <i>Internet of Things</i> (IoT) en la Banca.....	38
4.3 <i>Blockchain</i>	42
4.4 La Inteligencia Artificial.....	43
4.5 Impacto de la Inteligencia Artificial en la Banca.....	45
5. Conclusiones	48
6. Bibliografía	54

Índice de figuras

Figura 1. Historia de <i>Big Data</i>	8
Figura 2: “Esquema de los pasos del proceso de descubrimiento de conocimiento” ...	18

1. Introducción

1.1 Justificación de la importancia del tema

La era de *Big Data* ha venido con grandes oportunidades y desafíos, y todos los sujetos presentes en los mercados internacionales están sufriendo un desbordamiento de la cantidad de información que se genera a un ritmo impredecible. En 2015 Facebook tenía un promedio de 1.04 billones de usuarios activos al día, 934 millones de usuarios activos diarios que se conectaban desde sus dispositivos móviles, 205 billones de fotos subidas todos los días y cerca de unos 2.7 billones de *likes* y comentarios publicados al día (Bhadani y Jothimani, 2016). Estos datos recogen implícitamente un mayor detalle sobre los comportamientos de compra, las actividades de los clientes, el funcionamiento y el estado de las dinámicas sociales y culturales e incluso de los eventos que ocurren a nuestro alrededor. Cuando hablamos de *Big Data* nos referimos a toda información capturada por una organización y que es demasiado grande y compleja para ser analizada y procesada por los sistemas técnicos y de software tradicionales.

Por definición, el término *Big Data* se refiere “al conjunto de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento dificultan su captura, gestión, almacenaje y procesamiento mediante tecnologías o herramientas convencionales)” (Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad, 2019). En resumen, se trata de todo lo que engloba almacenar y procesar grandes volúmenes de datos estructurados, no estructurados o semi estructurados que explicaremos más adelante y que participan e influyen en la actividad diaria de la empresa.

El objetivo, por tanto, es ser capaz de analizar, integrar y hacer un uso correcto de estos volúmenes grandes de datos debido a que aportarán a toda organización ventajas comparativas significantes, como por ejemplo una mejor toma de decisiones basada en información y argumentos más sólidos. Por ello, ser capaz de entender qué hay detrás de ese gran volumen de datos se ha convertido en una tendencia emergente y en el objetivo final de una amplia gama de estudios.

El sistema financiero y los bancos en concreto se caracterizan por tener una cantidad de datos intensiva, y por ello, se han convertido en un campo de aplicación popular para

los investigadores de las técnicas del análisis de *Big Data*. La recopilación de información en la industria bancaria no estaba planeada, ya que se lleva haciendo desde hace tiempo, pero se ha materializado en ser una nueva línea de ingresos para los bancos y sus clientes se está ya haciendo mediante la extracción de valor de dicha información que se encuentra rápida y fácilmente a su alcance. Por todo esto, los bancos se han dado cuenta de que su mayor activo y, por tanto, fuente de ingresos es el conocimiento y no los activos financieros.

1.2 Objetivos

El objetivo general de este trabajo es llegar a entender cómo el análisis del *Big Data* ha transformado el valor de la información y las propuestas de negocio en el sistema financiero y en concreto en la industria bancaria. Buscaremos en un primer momento definir el concepto de *Big Data* mediante un análisis del recorrido histórico y de su evolución. Una vez comprendido el concepto de *Big Data* y su definición se buscará entender la forma en la que organizaciones pueden obtener, almacenar y procesar dichos datos para su consecuente aplicación a modelos de negocio. Además, explicaremos las distintas posibilidades y variantes en cuanto al uso que se puede hacer de dicha información para potenciar la creación de valor en las empresas.

Más adelante, nos centraremos en la aplicación del *Big Data* en la Banca comenzando por explicar por qué los bancos deciden y considerar relevante destinar recursos económicos al estudio de los datos almacenados en sus sistemas, así como las múltiples aplicaciones dentro del conjunto que engloba la banca. En este mismo apartado se buscará asimismo determinar y explicar los beneficios y potenciales riesgos que el sistema bancario ha sufrido mediante la aplicación de dicha ciencia.

Finalmente, se explorará las posibilidades y futuras opciones que ofrece el correcto análisis de los datos. Para ello, no solo se analizará la transformación que ha sufrido la ciencia y las técnicas del análisis del *Big Data*, sino se determinará también las posibles variantes que han surgido y que muchos expertos predicen que surgirán en el futuro.

1.3 Metodología

Con el fin de alcanzar los objetivos del estudio mencionados anteriormente, la metodología se basará en el estudio y la investigación exploratoria de técnicas cualitativas, más concretamente en una revisión de la bibliografía de fuentes primarias y secundarias.

En primer lugar, para definir las características principales que constituyen el concepto de *Big Data* se han utilizado fuentes primarias sacadas de bases de datos como Google Scholar, EBSCO o ResearchGate. De manera de encontrar la información adecuada se han utilizado palabras clave y conceptos clave como “Qué es *Big Data*”, “Definición *Big Data*”, “características *Big Data*”, “evolución *Big Data*”, “almacenamiento de la información”, “creación de valor a través del *Big Data*”, “*Big Data* y nuevos modelos de negocio”, “*Big Data* en el sector financiero”, “*Big Data* en la Banca”, etc. No se ha introducido ninguna fecha que delimite el espacio temporal de las búsquedas, aunque la mayoría de los documentos y de las fuentes analizadas se han basado principalmente en las dos primeras décadas del siglo XXI ya que es cuando el concepto y la ciencia detrás del *Big Data* ha cogido forma.

El análisis y la comprensión de la literatura se ha hecho mediante el análisis de bases de datos, artículos de periódicos científicos, artículos de investigación, trabajos de investigación, libros, artículos en periódicos, etc. Por un lado, fuentes periódicas, sitios web y publicaciones de muchas entidades bancarias han servido para definir los conceptos más recientes y más innovadores, como por ejemplo para definir los términos *Big Data* y su significado y para analizar sus potenciales usos y aplicaciones en la banca del futuro. Sin embargo, los artículos científicos, y los trabajos de investigación han permitido dar forma y definir más concretamente las partes en las que buscamos entender cómo el *Big Data* se puede aplicar a la banca, y qué medidas deberían los bancos de adoptar para poder convertirse y adaptarse en el posible “banco del futuro”. Muchos bancos y empresas dedicadas a la consultoría han analizado el futuro del *Big Data* y el futuro de las tecnologías y sus métodos de aplicación al modelo bancario.

Asimismo, los libros, reflexiones, trabajos de investigación analizados han permitido construir una imagen más amplia de lo que supone el *Big Data* en la Banca a día de hoy. Gracias a la lectura y al análisis de estos documentos se han podido identificar

las aplicaciones en las que ya se integran los modelos y las tecnologías y técnicas relacionadas con el *Big Data* y se ha podido dar más detalle a los aspectos positivos y negativos de la implementación de éstas.

2. ¿Qué es *Big Data*?

2.1 Definición y conceptos (3vs y 5vs)

Podríamos referirnos al término *Big Data* como “a la tendencia en el avance de la tecnología que ha abierto puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos que tomaría demasiado tiempo y sería muy costoso cargarlos a una base de datos relacional para su análisis” (Barranco Fragoso, 2012).

“El crecimiento en el volumen de datos generados por diferentes sistemas y actividades cotidianas en la sociedad ha forjado la necesidad de modificar, optimizar y generar métodos y modelos de almacenamiento que suplan las falencias de los sistemas de gestión de datos tradicionales” (Hernández-Leal, Duque-Méndez, y Moreno-Cadavid, 2017). Y por ello, dicho concepto aplica para toda aquella información que no puede llegar a ser analizada mediante la utilización de medios o procesos convencionales.

Entre las definiciones de *Big Data*, cabría también mencionar la definición del *National Institute of Standards and Technology* en Estados Unidos que lo define como un extenso conjunto de datos que requieren de una arquitectura especial para poder ser almacenados, manipulados y analizados eficientemente (Pritzker, 2015).

El término *Big Data* se utiliza para describir un gran volumen de datos, estructurados, no estructurados o semi estructurados, que están presentes en el día a día de un negocio. Sin embargo, lo que hace importante y relevante al *Big Data* no es la cantidad de los datos que podamos acumular, sino el uso que organizaciones y empresas hacen de dichos datos para sacarle provecho. En definitiva, el *Big Data* puede ser analizado para ayudar a empresas a tomar mejores decisiones y a aplicar mejores estrategias de negocio (SAS, 2019).

A pesar de que el término “*Big Data*” sea un término novedoso, las empresas y otros actores llevan recogiendo y almacenando enormes cantidades de datos desde hace bastante tiempo. El concepto empezó a usarse durante la década de los 2000 cuando Doug Laney (SAS, 2019) definió *Big Data* como las tres Vs: Volumen, Velocidad y Variedad.

Cuando hablamos de Volumen nos referimos a la cantidad de datos que son procesados, son, por tanto, los Petabytes de información que necesitan ser analizados y

diseccionados para que la información sea utilizable (What is Big Data?: The Three Vs, 2016). Las organizaciones recogen datos de muchas fuentes como pueden ser por ejemplo transacciones de negocios o información recogida en las redes sociales para almacenarla, procesarla y luego poder utilizarla. La Velocidad se refiere al tiempo que tarda la información, el *Big Data*, en ser procesada. Cuanto más se reduzca ese tiempo, mayor ventaja comparativa se puede obtener. Las empresas dependen de esta información, más específicamente, ciertas actividades empresariales necesitan que la información esté disponible en el momento adecuado para poder tomar las decisiones adecuadas (The five V's of big data, 2017). Finalmente, la última V se refiere a Variedad, que son los tipos de datos que el *Big Data* puede abarcar. Como hemos mencionado anteriormente, estos datos pueden ser estructurados o no estructurados y el análisis de *Big Data* puede incluir cualquier tipo de dato (Hadi, Shnain, Hadishaheed, y Ahmad, 2015). La importancia de estas fuentes de información varía en función de la naturaleza de la actividad empresarial (The five V's of big data, 2017).

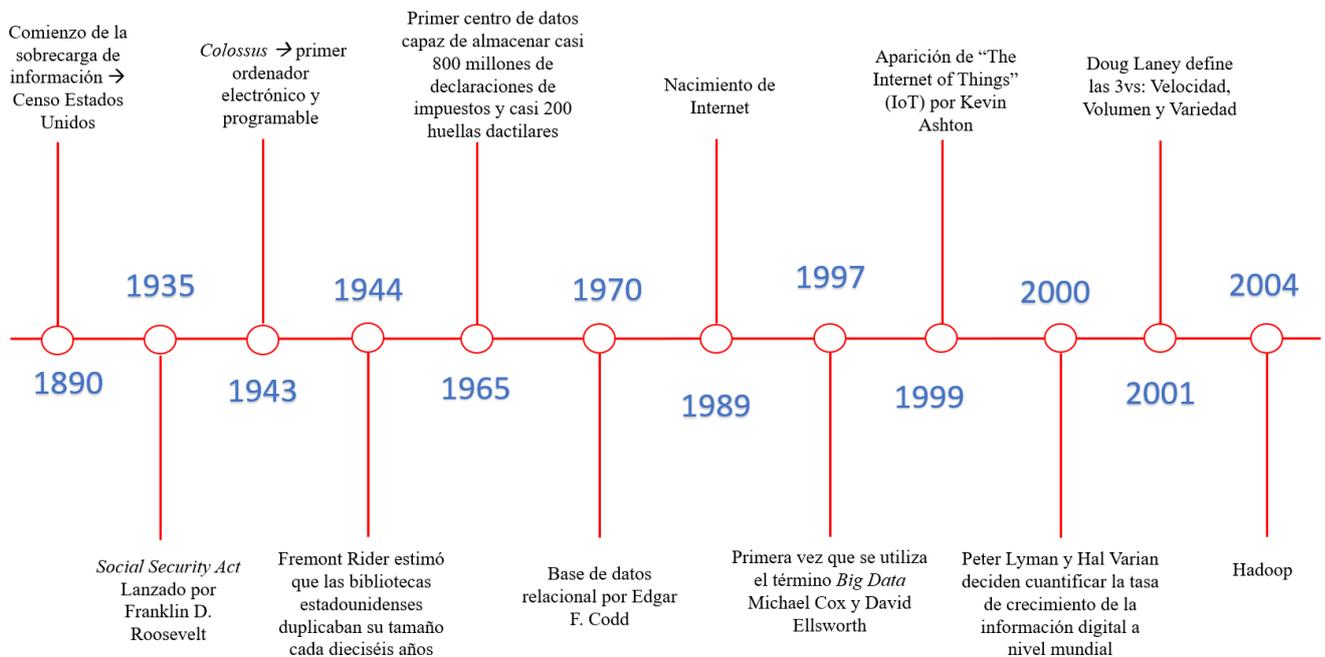
No obstante, IBM introduce otras dos V más: Veracidad y Valor. La primera se refiere al grado en el que un líder confía en la información que tiene para tomar una decisión (Hadi, Shnain, Hadishaheed, y Ahmad, 2015) ya que no todos los datos generados son correctos y por ello es importante estar seguro de la veracidad de la información que se obtiene. Esta variable se refiere, por tanto, al nivel de fiabilidad asociado a los datos. En cuanto al valor de la información podríamos definirlo como el valor añadido que los datos recogidos pueden aportar al proceso o a la actividad empresarial. Se trata de la parte más difícil y del componente más importante del *Big Data* ya que toda empresa buscará ser capaz de analizar, cruzar, transformar y extraer información fiable y de valor para tomar las decisiones correctas. (Hadi, Shnain, Hadishaheed, y Ahmad, 2015).

Cabe asimismo definir lo que son los datos estructurados, no estructurados y semi estructurados. Un conjunto de datos estructurados son aquellos datos que pueden utilizarse en su forma original para obtener resultados. Estos datos pueden ser ordenados (suelen aparecer con columnas y filas con títulos) y la mayoría de los sistemas informáticos están programados para generar este tipo de datos ya que facilita su procesamiento. Un ejemplo sería una base de datos de algún cliente (Maruti Techlabs, 2019). Por otro lado, los datos no estructurados carecen de formato o de alineación

predeterminada y adecuada y por ello estos conjuntos de datos aleatorios requieren de un mayor poder de procesamiento y de más tiempo para convertirlos en resultados tangibles. Y finalmente, los conjuntos de datos semi estructurados se encuentran a caballo entre los dos anteriores, de modo que pueden tener una estructura adecuada y, sin embargo, carecen de elementos definitorios para su clasificación y su procesamiento. Aquí hablaríamos de datos como RFID o XML (Maruti Techlabs, 2019).

2.2 Historia y evolución

Figura 1: Historia del *Big Data*



Fuente: Elaboración propia con fuentes: (Zarate Santovena, 2013) y (Marr, 2015).

Como podemos observar en el gráfico anterior, el punto de inflexión que marcó el comienzo de la historia del *Big Data* se atribuye al Censo realizado en Estados Unidos de 1890. Se trata del primer censo compilado utilizando unos métodos y una tecnología creada por Herman Hollerith. Hollerith decidió diseñar una máquina capaz de recopilar información más eficientemente mediante el uso de la detección eléctrica. Gracias a esta creación, el Censo americano de 1890 fue publicado después de seis semanas de estudio

mientras que el Censo de 1880 tardó más de ocho años en tabularse. (Zarate Santovena, 2013). Este invento no solo prevaleció durante unos cuantos años, sino que fue utilizado por IBM para la programación de sus primeros ordenadores. En 1935, como solución al boom demográfico sufrido en Estados Unidos después de la Primera Guerra Mundial, el presidente Franklin D. Roosevelt, lanza el “*Social Security Act*” (Zarate Santovena, 2013). IBM se encargaría entonces de mantener registros laborales de los 26 millones de trabajadores estadounidenses y de los 3 millones de empleadores de aquella época.

En 1943, se creó en Reino Unido, el primer ordenador electrónico programable llamado el *Colossus* para descifrar los códigos secretos utilizados por los Nazis durante la Segunda Guerra Mundial. Este ordenador tendría una significativa influencia en la creación de los primeros ordenadores, ya que confirmó que sería viable crear dispositivos informáticos de alta velocidad (Zarate Santovena, 2013).

En 1944, Fremont Rider (escritor y bibliotecario americano) publicó un trabajo llamado “*The Scholar and the Future of the Research Library*” en el que estimó que las bibliotecas estadounidenses duplicaban su tamaño cada dieciséis años, siendo esto una consecuencia inmediata del crecimiento del conocimiento. Teniendo en cuenta esto, Rider estimó que, si el método de almacenamiento de datos continuaría siendo en libros, para el 2040, la biblioteca de la Universidad de Yale tendría aproximadamente 200 millones de ejemplares, ocupando más de 9.000 km y que requeriría más de seis mil trabajadores (Zarate Santovena, 2013).

Hasta este momento, el único desafío de los datos parecía estar relacionado con el volumen, sin embargo, en durante la década de 1960, un nuevo componente clave para el *Big Data* vio la luz: la Velocidad. Concretamente en 1961, Derek Price publica “*Science Since Babylon*” en la que traza el crecimiento del conocimiento científico mediante el análisis del crecimiento del número de revistas y documentos científicos. Concluye que el número de revistas ha crecido exponencialmente, multiplicándose por dos cada quince años y multiplicándose por diez cada medio siglo (Zarate Santovena, 2013).

Debido a que la importancia y el crecimiento del volumen de la información no se daba únicamente en el sector científico, muchas organizaciones empezaron a desarrollar sistemas para automatizar sus inventarios En vez de limitar la cantidad de datos que se pueden almacenar, el objetivo era ser capaz de desarrollar una nueva tecnología que

permita dar un uso más eficiente al espacio disponible. En 1965 el Gobierno estadounidense decide construir el primer centro de datos para almacenar más de 742 millones de declaraciones de impuestos y 175 millones de huellas dactilares convirtiendo todos esos registros en cintas magnéticas para ordenador que podían así ser almacenadas en un único lugar. El proyecto fue finalmente abandonado, pero se aceptó generalmente, que dicho centro fue el comienzo de la era del almacenamiento electrónico de datos (Zarate Santovena, 2013).

En 1970 el matemático graduado de Oxford, Edgar F. Codd, presenta un modelo para la creación de una “base de datos relacional” (Marr, 2015). Este modelo es disruptivo en el sentido de que hasta el momento se necesitaba tener avanzados conocimientos informáticos para ser capaz de recuperar la información. Además, se sigue utilizando hoy en día para la mayoría de las transacciones de datos cotidianas (como por ejemplo para el uso de una tarjeta de crédito). (Marr, 2015).

Sin embargo, un evento de gran importancia es el nacimiento del Internet en 1989. El informático británico Tim Berners-Lee inventa “the World Wide Web”, una red de datos interconectada a nivel mundial, accesible para cualquier persona desde cualquier lugar del mundo (Marr, 2015). En 1997, se inaugura la revista “*Knowledge Discovery and Data Mining*” que se centra en la teoría, en las técnicas y en la práctica necesaria para poder extraer información de grandes bases de datos. Es en este mismo año en el que se utiliza por primera vez el término “*Big Data*” por Michael Cox y David Ellsworth (Zarate Santovena, 2013). Y en 1999, se da posiblemente el primer uso al término “*Internet of Things*” (Marr, 2015) o IoT por Kevin Ashton para describir el creciente número de dispositivos conectados online y el potencial que tienen para conectarse entre sí, sin la necesidad de que haya algún humano como mediador.

A principios del siglo XXI, más concretamente en el año 2000 es cuando por primera vez dos economistas intentan no solo cuantificar la cantidad de información digital en el mundo sino también, intentan calcular su tasa de crecimiento. Peter Lyman y Hal Varian (economista jefe de Google) concluyeron que la producción anual de impresión, de películas, y de contenido óptico requería aproximadamente 1.500 millones de gigabytes de almacenamiento. Y en el año 2001, el analista Doug Laney define las tres

características comúnmente aceptadas para describir al *Big Data* llamadas las 3Vs: Velocidad, Volumen y Variedad.

En 2004 nace Hadoop, uno de los principales elementos detrás de la revolución del *Big Data* y fue creado específicamente para el almacenamiento y el análisis de grandes volúmenes de datos. Hadoop tiene cuatro características importantes: Escalabilidad, Bajo costo, Flexibilidad y Tolerancia a fallos. La escalabilidad se refiere a que Hadoop es capaz de hacer crecer su sistema para que procese más datos y debido a que la estructura de código abierto es gratuita, se caracteriza por tener un bajo costo. Por otro lado, es flexible ya que “a diferencia de las bases de datos relacionales, no tiene que procesar previamente los datos antes de almacenarlos y por ello puede almacenar tantos datos como desee y decidir cómo utilizarlos más tarde” (Hadoop ¿Qué es y por qué es importante?, 2019). Y además puede gestionar todo tipo de datos, incluso los no estructurados (como los vídeos, o imágenes). Y finalmente, “el procesamiento de datos y aplicaciones está protegido contra fallos del hardware” (Hadoop ¿Qué es y por qué es importante?, 2019).

Pero quizás el mayor hito de la evolución tecnológica y del *Big Data* puede haber sido *Watson*, un sistema informático de inteligencia artificial capaz de contestar a preguntas. *Watson* fue desarrollado como parte del proyecto de DeepWA de IBM por un equipo liderado por David Ferruci y fue explícitamente creada para contestar a preguntas de un concurso televisivo (Zarate Santovena, 2013). Se trata de una creación sin precedente ya que depende de capacidades humanas que van más allá de la capacidad tradicional de un ordenador. Entre estas capacidades podemos encontrar entre otras cosas por ejemplo la habilidad de dar respuestas extremadamente rápidas. *Watson* tenía acceso a 200 millones de páginas de contenido estructurado y no estructurado y consumía cuatro terabytes (1.024 gigabytes) de almacenaje de disco (Zarate Santovena, 2013).

2.3 Cómo obtener los datos

Datos sobre cada uno de nosotros se recogen a diario a través de acciones cotidianas y que realizamos con regularidad como por ejemplo al utilizar una tarjeta de crédito, al completar una encuesta, al irnos de vacaciones o al enviar un correo electrónico (Ferris, Moore, Pohle, y Srivastava, 2014). Pero, además, todas aquellas personas que utilizan las redes sociales permiten que se transmita cualquier tipo de información adicional sobre

ellos y sobre aquellos con los que se comunican. Mientras que algunos de estos datos requieren de nuestro permiso para ser utilizados o incluso recogidos por terceros, otros datos informativos pueden llegar a recogerse sin que el individuo lo sepa. En un primer momento abordaremos y explicaremos las cuatro maneras de analizar el comportamiento de un consumidor y a continuación explicaremos y nombraremos algunos de los métodos de recolección de información.

En un primer lugar podríamos llevar a cabo un análisis descriptivo en el que incluiríamos datos demográficos (género, edad, geografía, ingresos económicos) pero también datos sobre las preferencias de un consumidor (Sauro, 2014) (categoría de producto, o tipo de tecnología). Un análisis del comportamiento nos llevaría a obtener patrones de comportamiento sobre los consumidores (Sauro, 2014) (a qué horas del día busca productos tecnológicos y cuándo se interesa y busca prendas de ropa). Además, se pueden realizar análisis interactivos que miden las actividades de navegación de cada usuario en los sitios web. Y finalmente, el análisis de actitud nos aportará información sobre los gustos, las preferencias, las opiniones y los sentimientos (Sauro, 2014).

Existen varios métodos que se pueden usar para recoger y obtener datos, como por ejemplo las encuestas (dónde se busca obtener información de determinados sujetos de estudio y pueden coger la forma de entrevista o de cuestionario) u observaciones (siendo estos registros visuales de lo que ocurre en una situación real). Sin embargo, en la actualidad, gracias al uso de internet y de las redes sociales, existen otros métodos de obtención de datos adicionales. La recogida de información a través de las redes sociales se diferencia de los mecanismos tradicionales en un primer lugar por su naturaleza, y es que los investigadores no intervienen activamente en el proceso y eso hace que el mecanismo de obtención y el estudio de la información sea más objetivo (Liang y Zhu, 2017). Sin embargo, una desventaja es la dificultad de conocer las variables demográficas y psicológicas de los usuarios. Se trata de un proceso automático que permite obtener información global en tiempo real.

Los actores que son capaces de recolectar una mayor cantidad de información sobre sus consumidores son Google, Yahoo! o Bing, mundialmente conocidos por sus buscadores de información. A pesar de que seguramente sus buscadores sean la mayor fuente de recolección de datos para estas empresas, todas ellas tienen además otro tipo de

herramientas (Ferris, Moore, Pohle, y Srivastava, 2014). Un ejemplo claro sería Google, que además de su buscador tiene otro tipo de servicios como Gmail, que le permite recolectar más información y así configurar una estrategia de marketing hecha a medida para cada persona. Como bien sabemos a medida que Google, en su amplia gama de servicios reconoce palabras que escribimos, objetos que deseamos, sitios a los que queremos ir, hará que nos aparezcan progresivamente anuncios relacionados con nuestro perfil, gustos y preferencias. En este tipo de situaciones, el usuario normalmente no paga ninguna cuota para utilizar el servicio, simplemente introduce sus datos voluntariamente y usa una página web que es propiedad de un tercer actor. Debido a que el sujeto introduce voluntariamente sus datos en este tipo de páginas, el que verdaderamente es dueño de toda aquella información es el propietario de la página web.

Otro método de recogida de datos se hace a través de proveedores de marketing, como IBM, Acxiom o Equifax (Ferris, Moore, Pohle, y Srivastava, 2014). Estas empresas se dedican a colaborar con otras empresas y a recoger datos de clientes que luego meten en bases de datos que venden. Por lo general, estas empresas son capaces de conseguir información de gran alcance, como por ejemplo información sobre el comportamiento de compra de consumidores, o datos salariales de estos. Cada vez que un consumidor compra algo con su tarjeta de crédito, todos los datos de la transacción se recogen. Una vez recogida dicha información se agrega y se consolida para poder vendérsela a algún tercero, que, por tanto, tendrá acceso a información de consumidores sobre el tipo de productos que se compran, la localización de dichas compras, etc. En los últimos años, actores como empresas dedicadas a proveer servicios financieros o incluso empresas de seguros de vida que buscan obtener este tipo de información para proveer servicios más especializados y rentables.

Por supuesto las bases de datos tradicionales, también conocidas como repositorios de datos, constituyen una fuente importante de recolección y recogida de información y normalmente se especializan por industria, sector y geografía. Y finalmente, mediante el uso de aplicaciones del móvil, los propietarios de estas también son capaces de saber información muy detallada y muy personal sobre sus clientes. Un ejemplo son las aplicaciones para hacer deporte, o las aplicaciones que miden los kilómetros andados en un día, nos dan información sobre el estilo de vida de una persona (Ferris, Moore, Pohle, y Srivastava, 2014).

2.4 Almacenamiento de la información

El objetivo final de la implementación y del uso del *Big Data* es ser capaz de integrarlo a una empresa, una actividad o un modelo de negocio concreto. Sin embargo, “uno de los aspectos importantes a considerar en un sistema para la administración de datos es el almacenamiento; dado que de esto dependen muchos factores como el acceso, la disponibilidad, la escalabilidad, la facilidad de recuperación, estructuración de consultas, tiempos de respuesta, entre otras.” (Hernández Leal, 2016).

El almacenamiento de *Big Data* busca principalmente almacenar y gestionar datos de manera escalable. De esta manera el almacenamiento satisface las necesidades de aplicaciones que necesitan acceder a esta información, es decir, el almacenamiento permite encontrar respuestas a preguntas complejas. El sistema ideal de almacenamiento de grandes datos debería permitir el almacenamiento de una cantidad prácticamente ilimitada de datos, ser capaz de tratar eficiente y flexiblemente una gama de diferentes modelos de datos, poder soportar tanto los datos estructurados, como los no estructurados, y finalmente, por razones de seguridad, debería trabajar únicamente en datos cifrados. (Strohbach, Daubert, Ravkin, y Lischka, 2016). Entre los métodos de almacenamiento para grandes volúmenes de datos encontramos desde modelos más tradicionales que son las conocidas bases de datos relacionales, hasta nuevos modelos que son capaces de almacenar también los datos no estructurados y los semiestructurados.

- **Bases de datos relacionales**

Una base de datos relacional trata con un conjunto de tablas y cumple con el modelo relacional (basado en el uso de relaciones, siendo esta un vínculo entre dos entidades de la base de datos (Herranz Gómez, 2014)). Fueron creadas por Edgar Frank Cobb en 1970y tienen varias características: están compuestas de varias tablas o relaciones; no puede haber dos tablas que tengan el mismo nombre ni registro (Hernández Leal, 2016)y, las tablas se relacionan entre sí mediante claves primarias y claves foráneas. Este método sigue siendo el más usado hoy en día para cualquier tipo de aplicación informática, como por ejemplo para aplicaciones transaccionales hechas por muchos usuarios, éstas les permiten actualizar las bases de datos continuamente y requieren de consultas poco completas. (Herranz Gómez, 2014).

- ***Datawarehouse***

Un *Datawarehouse* es como un almacén unificado donde se recoge información histórica de una o más fuentes distintas, que luego procesa a gran velocidad. Se construyen mediante un proceso en el que limpia la información, se integra, se transforma, se agrega la información y se actualiza periódicamente. Más concretamente, se caracteriza por ser integrado, temático, histórico y no volátil.

Integrado quiere decir que los datos que se van a almacenar se tienen que integrar consistentemente: estos datos entonces se tendrán que depurar y modificar para que se estructuren en base a un esquema con un gran nivel de detalle, eliminando así las inconsistencias. La característica temática se refiere a que los datos deben clasificarse por temas, para que el usuario final sea capaz de entenderlo mejor y de acceder más rápida y fácilmente a dicha información (Datawarehouse, 2019). Además, este método sirve para realizar análisis de tendencias y comparaciones entre datos, y por ello, cada variable tendrá los distintos valores que tome con el paso del tiempo. Finalmente, que sea no volátil quiere decir que “al contrario como pasa en los sistemas transaccionales tradicionales (donde se inserta y modifica información de forma constante), en un *datawarehouse* los datos se cargan y acceden generalmente de forma masiva sin ser modificados” (Hernández Leal, 2016).

- **NoSQL**

Finalmente, el último método que vamos a estudiar son las bases de datos NoSQL, que significa “*Not Only SQL*”. Se trata de un grupo de bases de datos no relacionales que analizan y contemplan datos no estructurados o semiestructurados. El aparecimiento de este tipo de bases de datos se da con la llegada de aplicaciones como Facebook, Twitter o Youtube porque hasta entonces pocas empresas subían datos a la red, pero con la aparición de estas, la cantidad de datos empezó a crecer exponencialmente.

Con el crecimiento de la información, empiezan a surgir problemas de gestión en las bases de datos tradicionales, llamadas relacionales. Una primera solución fue aumentar el número de máquinas de almacenamiento para las bases relacionales, pero pronto los especialistas se dieron cuenta de su ineficiencia y de los altos costes que esta solución

requería. Con el tiempo, la solución a la que se llegó fueron las bases de datos NoSQL. “Por lo tanto, hablar de bases de datos NoSQL es hablar de estructuras que nos permiten almacenar información en aquellas situaciones en las que las bases de datos relacionales generan ciertos problemas debido principalmente a problemas de escalabilidad y rendimiento de las bases de datos relacionales donde se dan cita a miles de usuarios concurrentes con millones de consultas distintas” (Acens).

Entre las ventajas principales de este sistema frente a las bases de datos relacionales podemos recalcar el hecho de que se ejecutan en máquinas que no requieren de muchos recursos y por ello, el coste es menor; por otro lado, permiten que haya una escalabilidad horizontal; es capaz también de manejar una mayor cantidad de datos y son fáciles de utilizar. Sin embargo, entre las principales desventajas encontramos “que no cuentan con un único modelo de datos a nivel de sistema, en cuanto a la estructura se tiene poca estandarización de interfaces para servicios, tampoco tiene una semántica estándar y esto hace que traiga consigo problemas de interoperabilidad” (Hernández Leal, 2016).

2.5 Cómo analizar-procesar los datos

Cuando hablamos del análisis de *Big Data*, existen ciertos problemas que pueden obstaculizar las fases del proceso de creación de valor. Entre estos problemas podemos encontrar la escalabilidad, la complejidad, la privacidad y la heterogeneidad, entre otros. La realidad es que muchos datos hoy en día no son estructurados, por ejemplo, un *tweet* son partes de texto que están poco estructuradas, mientras que una imagen o un vídeo están estructuradas para almacenarse y reproducirse, pero no para buscarse por contenido semántico. Lo que verdaderamente se busca, es ser capaz de transformar dicho contenido en un formato estructurado que pueda analizarse luego. El valor de los datos aumenta cuando se puede vincular con otros datos de información, por lo que la integración de datos es un paso importante para la creación de valor (Duggal y Paul, 2013)

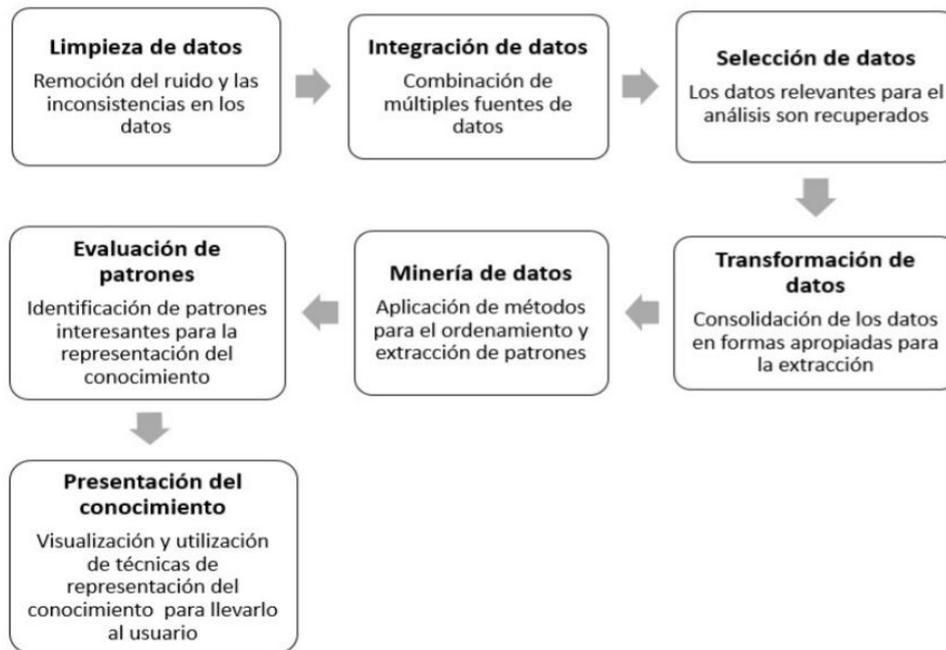
El *Big Data* permite proporcionar análisis sobre información, sobre conjuntos de datos. Sin embargo, resulta importante resaltar que realizar un análisis no es lo mismo que realizar un reporte. El *Reporting*, o la presentación de informes, es el proceso de organización de los datos en resúmenes informativos que permiten supervisar el desempeño de las distintas áreas de un negocio. No obstante, el análisis va un paso más

allá y es considerado como el proceso de exploración de datos o de informes con el objetivo de extraer conclusiones significativas que ayudan a comprender cómo mejorar el rendimiento empresarial. En definitiva, un reporte traduce los datos en información, y ayuda a las empresas a poder controlar sus negocios online y a ser capaz de darse cuenta de cuándo algo no esté yendo como debería. Y el análisis, transforma los datos y la información que aporta el reporte en conocimiento y luego en recomendaciones.

Una técnica para el análisis de información es la Minería de datos que permite extraer conocimiento útil de una base de datos. Es muy difícil encontrar el valor directamente de cada dato, y por eso la minería de datos necesita de un preprocesamiento y de unas técnicas analíticas para encontrar el valor, para encontrar dichos conocimientos útiles. Debido al crecimiento de la información, las empresas tuvieron que permitir que se implementara el uso de la minería de datos en sus distintas áreas del negocio, “puesto que de esta manera se permite el descubrimiento automático o semiautomático de información relevante a partir de estos cúmulos de datos” (Hernández Leal, 2016).

De hecho, la técnica de minería de datos está relacionada con la inteligencia artificial y el *Machine Learning* (Adam, Majid, y Zain, 2014). La escala de la gestión de datos del *Data Mining* y del *Big Data* es diferente debido a las diferencias en el tamaño (Adam, Majid, y Zain, 2014). Sin embargo, el método básico para la extracción de valor es muy similar. Cuando hablamos de *Data Mining*, el proceso de extracción de información útil necesita que los datos se limpien, se integren, se seleccionen, se transformen, se extraigan, se minen los datos, se evalúen y finalmente, se presente el conocimiento. Y el *Big Data* llegó para proponer soluciones a ciertas cuestiones, como por ejemplo el manejo de distintos tipos de datos.

Figura 2: “Esquema de los pasos del proceso de descubrimiento de conocimiento”



Fuente: (Hernández Leal, 2016).

Por otro lado, el análisis del *Big Data* es diferente del análisis de datos tradicional en que los métodos utilizados para analizar datos tradicionalmente son incompatibles con el gran aumento en el volumen de la información. Simplemente, era necesario que se crearan nuevas tecnologías innovadoras, y Google encontró la solución mediante el uso de un modelo de procesamiento conocido como “*MapReduce*” (Adam, Majid, y Zain, 2014). Entre las múltiples soluciones, la más conocida y más utilizada es el procesador de infraestructura *Hadoop*, un proyecto basado en Google *MapReduce* y en el sistema de archivos de Google (Adam, Majid, y Zain, 2014).

Hadoop se presenta como una infraestructura de sistema distribuido que fue investigado y desarrollado por la Fundación Apache. Entre las varias aplicaciones de Hadoop, cabe destacar particularmente la aplicación al buscador Yahoo! El buscador Yahoo! es el líder en investigación y aplicaciones tecnológicas del sistema Hadoop. La empresa aplica Hadoop a varios productos entre los cuales encontramos el análisis de datos, la optimización de contenidos, un sistema de correo electrónico antispam y finalmente, la optimización de la publicidad. Hadoop también se ha usado para determinar los intereses de los usuarios, en el buscador y en dónde poner publicidad. (Wu, 2015).

2.6 Cómo usar los datos para crear valor

Lo único que nos queda por analizar es cómo se pueden utilizar los datos para conseguir crear valor. Las organizaciones son uno de los principales actores hoy en día dependientes de la información de calidad y de los datos que disponen ya que de ellos depende el éxito futuro de su negocio. Vivimos en un mundo en el cual se generan enormes cantidades de datos en todas partes: Google solamente recibe 2 millones de consultas de búsqueda cada minuto y los usuarios de Facebook publican alrededor de 700 mil piezas de contenido al mismo tiempo (Kubina, Varmus, y Kubinova, 2015). La aparición del término *Big Data* cambió el significado de los datos, permitiéndonos sacar información de calidad a esos datos e implementarlos dentro de las empresas con el objetivo de obtener mayores beneficios. Esto es posible, en gran parte, por la aparición de las Redes Sociales (Facebook, LinkedIn, Twitter, Instagram...). Estas plataformas tecnológicas han revolucionado la forma de conectarse con otra gente, no solo permiten que nos conectemos constantemente los unos con los otros en todas partes, sino que también aportan información sobre tendencias.

Sin embargo, hoy en día las redes sociales no lo son todo ya que las personas utilizan muchos canales de comunicación (correo electrónico, SMS, etc.) que dejan lo que se conoce como la huella digital. La huella digital se define como “el rastro que deja una persona al navegar en internet y al actuar con el ciberespacio, desde las compras que se realizan en una tienda online hasta los comentarios y reacciones que se hacen en las redes sociales y que permiten dibujar un retrato digital” (Ramírez, 2018). Este rastro tiene un potencial de negocio para las empresas porque les permite conocer mejor a su cliente a través de métodos analíticos avanzados (como el Hadoop) que tratan datos no estructurados. El *Big Data* es importante para las empresas porque les permite, como ya hemos mencionado, sacar el valor de toda la información que hay en el mundo. Las empresas y organizaciones almacenaban datos que eran parte de cada transacción que hacían. Está información se usaba para el seguimiento o para prever el futuro. Pero hoy en día con la explosión de información, es posible que se recoja información de cada consumidor que visita una página web y, por tanto, saber lo que piensa el cliente en cada momento.

Es más, de acuerdo al McKinsey Global Institute, el *Big Data* es capaz de crear valor de distintas formas (Manyika, y otros, 2011).

- **Creando transparencia**

Cuando se permite que los accionistas y/o las partes interesadas tengan acceso a *Big Data* en periodos de tiempo concretos, las empresas son capaces de crear valor. Por ejemplo, en el sector público, cuando los distintos departamentos pueden acceder a datos pertinentes o que necesitan puede llegar a reducir notablemente el tiempo de búsqueda de estos y de procesamiento (Manyika, y otros, 2011).

- **El *Big Data* permite descubrir necesidades y mejora el rendimiento**

A medida que las empresas y organizaciones almacenan datos de forma digital, también recopilan información sobre el rendimiento de éstas de una forma más detallada y más precisa. Además, con el uso del *Big Data* podemos analizar si algunos eventos dentro de las organizaciones son debidos a eventos naturales o se generan por otros motivos, y así, los líderes podrán manejar y mejorar su rendimiento. (Manyika, y otros, 2011).

- **Segmentar poblaciones para proveer servicios diferenciados**

Como en Marketing, el *Big Data* permite a las organizaciones crear nichos de mercado muy específicos de manera que luego puedan crear productos y servicios a medida que satisfagan sus necesidades. Todas las empresas están empezando a usar el *Big Data* para ser capaz de tener técnicas cada vez más sofisticadas que permiten captar (en tiempo real) clientes. (Manyika, y otros, 2011).

- **Ayudan a la toma de decisiones**

El análisis avanzado de la información puede mejorar substancialmente la toma de decisiones, minimizar los riesgos y descubrir ideas valiosas que, sin esto, permanecerían ocultas (Manyika, y otros, 2011). Cuando una empresa analiza enormes conjuntos de datos utilizando técnicas y tecnologías del *Big Data*, en vez de muestras más pequeñas e insignificantes, puede ocurrir que las decisiones queden respaldadas y argumentadas por las conclusiones sacadas de dichos análisis.

- **Creación de nuevos modelos de negocio, productos y servicios**

El análisis del *Big Data* va a permitir a empresas descubrir nuevos modelos de negocio o productos para desarrollar. Las empresas están usando los datos obtenidos de productos actuales, que ya están en el mercado, para mejorar la generación futura de productos e incluso para crear respuestas y ofertas innovadora en los servicios de post venta. Es más, la “aparición de datos de localización en tiempo real ha creado un conjunto totalmente nuevo de los servicios basados en la localización, desde la navegación hasta los precios” (Manyika, y otros, 2011). Por ejemplo, las aseguradoras ahora ofrecen pólizas distintas en función de cómo conduzca cada persona su coche, entre otras cosas. (Manyika, y otros, 2011).

3. *Big Data* en Banca

La aparición e impacto del *Big Data* en las economías del mundo ha hecho que se cree un considerable interés en la industria de las tecnologías de la información. Debido a que cada vez tenemos más capacidad de generar, recopilar y almacenar datos digitales a gran escala, el concepto *Big Data* se ha hecho realidad. Gracias a que existe un mayor número de personas que cada vez están más conectadas, se generan constantemente datos por medio de transacciones online, de interacciones en redes sociales, por emails, búsquedas, registros, o incluso gracias a satélites de posicionamiento a nivel global. En total, la cantidad de datos que se producen al día ya superan los 2,5 Exabytes (Al Hujran y Al-Debei, 2015).

Como ya hemos mencionado en apartados anteriores, en el mundo de los negocios, las empresas están aprovechando las tecnologías de *Big Data* para tomar decisiones de una manera más efectiva y para mejorar sus ingresos por medio de un mayor impacto comercial. Gracias a estas tecnologías, las empresas serán capaces de conocer mucho mejor a sus clientes y, por tanto, de mejorar y personalizar sus productos y ofertas. El McKinsey Global Institute realizó estudios de *Big Data* en cinco áreas: la sanidad en los Estados Unidos, el sector público en Europa, el comercio en Estados Unidos y el comercio de fabricación a nivel mundial (Manyika, y otros, 2011). Su principal conclusión fue que el *Big Data* puede generar valor en cada uno de los ejemplos. El comerciante o fabricante minorista que utiliza *Big Data* puede aumentar su margen operativo en más de un 60%; si la sanidad estadounidense usara *Big Data* de manera creativa y eficaz, podrían generar más de 300 billones de dólares americanos en valor cada año, lo cual se usaría para reducir la deuda americana en un 8%; y en el sector público europeo, podrían ahorrarse más de 100 billones de dólares americanos en mejoras de eficiencia operacional gracias al *Big Data* (Manyika, y otros, 2011). Es más, los actores que definirán las reglas del juego en el escenario internacional son aquellos que tengan en su propiedad información clave y de extrema importancia.

Como hemos podido ver, la era del *Big Data* trae consigo muchas oportunidades y desafíos. La cantidad de información que se está generando es cada vez mayor, entre las empresas y los estudios ya no solo sirve con almacenar la información, sino que la actual tendencia y objetivo final de todo sujeto es ser capaz de entender y extraer el valor a los

datos. Por ello, para solucionar este problema ha aparecido la técnica de *Data Mining*, capaz de revelar la información oculta en los macrodatos (o *Big Data*) (Hassani, Huang, y Silva, 2018). Toda empresa, incluidos los bancos, han reconocido que la mayor ventaja comparativa para una empresa es el conocimiento, el poder de tener información de valor en sus manos. En el caso del sector bancario, gracias al crecimiento de la banca online y al éxito de la banca móvil, los bancos son capaces de obtener cantidades cada vez mayores de información sobre sus clientes, en tiempo real. Gracias al uso de las tarjetas de crédito, los bancos son capaces de saber dónde están sus clientes, sus gustos, sus preferencias, el salario que cobran, dónde viven, dónde pasan sus vacaciones, etc.

El continuo desarrollo y avance de las tecnologías hacen que cada vez haya una mayor disponibilidad de información y por ello, ser capaz de dominar las técnicas de análisis de *Big Data* se ha convertido en uno de los principales retos y objetivos del sector bancario. (Hassani, Huang, y Silva, 2018).

3.1 Motivos para la utilización de *Big Data* en la Banca

La revolución en las tecnologías de las comunicaciones ha provocado el imparable crecimiento que estamos viviendo de la cantidad de datos y de información producida en tiempo real. Todo negocio se ha visto en la necesidad de incorporar estos nuevos cambios en sus sistemas para tener mejores perspectivas de futuro. Un claro ejemplo es el modelo de negocio de Amazon en lo relacionado a la compra y venta de libros.

Hasta principios del siglo XXI, Amazon recomendaba libros, a través de una plataforma llamada *Amazon Voice*, que podían resultar de interés para sus clientes a través del conocimiento de múltiples expertos del mundo editorial, tanto conocidos y renombrados críticos como editores. Sin embargo, durante la década de los años noventa, Amazon decidió que quería invertir en la búsqueda de un proceso automatizado que realizase dicha tarea y Linden creó el conocido “filtro colaborativo *item-by-item*” (García Montalvo, 2014). Este algoritmo construye una lista de recomendaciones de compra, basada en productos que el cliente tiende a mirar repetidas veces o en productos similares a los que ya ha adquirido en el pasado. Este método buscaba desarrollar una capacidad predictiva y no explicativa, es decir, no necesita saber el porqué de las decisiones de compra de los clientes, sino que únicamente quiere predecir lo que le podría interesar en

base a lo que ya ha comprado que sea similar. Este método es importante ya que se ha integrado en muchas empresas a nivel mundial (*Netflix*, por ejemplo) y constituye la razón detrás de un tercio de las ventas de Amazon. Se trata también de la principal herramienta que produce una mayor satisfacción en los clientes, ya que gracias a ella pueden descubrir todo tipo de productos que puedan llegarles a interesar o necesitar (García Montalvo, 2014).

En base al algoritmo anteriormente mencionado es interesante ya que el objetivo último que persigue es ser capaz de ofrecer el mejor servicio posible mediante un grado mayor de personalización para cada cliente en concreto. ¿Podría el sistema financiero y en concreto el sistema bancario desarrollar un mecanismo parecido que responda a los intereses concretos de sus clientes? Según García Montalvo (García Montalvo, 2014), los bancos tienen que tener su foco en los segmentos poblacionales con dificultades de acceso o que directamente no tienen posibilidad de acceso a los servicios bancarios. ¿Cómo pueden conseguir esto? Gracias a las tecnologías centradas en el análisis de datos, los bancos deben crear fichas que analicen los distintos perfiles de dichos clientes. Y en función del perfil de riesgo y de la capacidad de pago, entre otras cosas, ofrecerles productos adecuados que tengan costes muy bajos. García Montalvo en su publicación también estima que en Estados Unidos existen alrededor de 60 millones de personas que no pueden acceder a servicios bancarios por el mero hecho de que no tienen un historial crediticio.

Por otro lado, en el mundo financiero existen una serie de factores que han motivado a todos los bancos a recoger, almacenar y analizar volúmenes de datos masivos (también conocido como en análisis del *Big Data*). Un cliente ya no necesita ir físicamente a una sucursal bancaria para hacer un depósito o para ejecutar una transacción bancaria; compradores y vendedores de acciones ejecutan sus roles conectados a la red por medio de ordenadores en tiempo real; las declaraciones de impuestos ya se realizan por los mismos individuos interesados en internet, sin la necesidad de acudir a un experto de impuestos para que le ayude a hacerla. (Gutierrez, 2018). El crecimiento de los negocios que están desarrollando plataformas de *e-commerce*, y en el caso de los bancos el crecimiento de la banca online ha permitido que los clientes puedan ya manejar autónomamente sus transacciones bancarias.

La forma de hacer negocio y de establecer relaciones con proveedores y clientes ha cambiado completamente. Los clientes no necesitan estar en contacto físico con el vendedor, por tanto, dicho vendedor no tiene por qué ser ni siquiera local o nacional, los compradores elegirán la opción más rentable y que más satisfaga sus necesidades, sea de donde sea. Por ello, fomentar la lealtad de un cliente hacia una marca se ha convertido en una tarea más complicada ya que es más difícil saber sus sentimientos o sus opciones. Sin embargo, otras oportunidades han aparecido gracias al uso de Internet: las empresas son capaces de capturar y analizar las acciones hechas por sus clientes con el fin de entender sus comportamientos y preferencias. (Gutierrez, 2018).

Gracias al desarrollo de las tecnologías, de las comunicaciones, y de las plataformas online, los usuarios tienen a su alcance un mayor abanico más amplio de posibilidades. Gracias al email y a las redes sociales los usuarios y las personas pueden estar conectadas con cualquier persona en cualquier parte del mundo instantáneamente. En el sistema bancario, los usuarios de las plataformas de la banca online tienen un mayor acceso hoy en día a la ejecución de transacciones financieras lo que ha llevado a que haya un crecimiento de la actividad bancaria y a que está busque expandirse hacia nuevos mercados. (Gutierrez, 2018). Por tanto, una mayor facilidad de acceso y una simplificación de las operaciones ha llevado a un aumento en el volumen de actividad de las entidades bancarias y en consecuencia, a un aumento de la cantidad de datos disponibles para analizar y utilizar.

El crecimiento en el volumen de la información ahora disponible ha traído consigo la aparición de nuevas fuentes de datos que son más complejas y difíciles de analizar que las tradicionales. Se trata de los datos no estructurados provenientes, mayoritariamente, de las redes sociales, que junto al resto de información que ya tenemos (transacciones individuales financieras) pueden llegar a crear una imagen agregada de las personas, de las organizaciones y del funcionamiento de los mercados (Gutierrez, 2018). La agregación de este tipo de información, junto con datos históricos permitirá alimentar modelos predictivos, previsiones y posibles impactos en la actividad comercial. Además, con la creación y el desarrollo de la banca online los bancos se han vuelto más competitivos y buscan captar cualquier oportunidad (a través de la creación de valor gracias al análisis de los datos) para vender y aumentar su clientela y sus ventas. Debido a que las tecnologías existentes (las tradicionales bases de datos relacionales) no son

capaces de procesar estas nuevas fuentes de información, la creación y adaptabilidad a las tecnologías *Big Data* se ha vuelto imprescindible.

Finalmente, existe una nueva tendencia hacia el cumplimiento de las normativas mediante la aplicación e implementación de leyes reglamentarias más rigurosas. Estas normas y requerimientos se han intensificado hacia áreas como la gobernanza de una empresa con el fin de fomentar la transparencia en las organizaciones a nivel mundial. Todas las industrias se han visto obligadas a cumplir con dichos requisitos, y el sistema financiero no ha sido menos. (Gutierrez, 2018). Además, nos encontramos actualmente en un periodo de bajos tipos de interés impuestos por los bancos centrales de manera a estimular la economía. Se impusieron estas medidas con la crisis económica del año 2009, pero después de casi diez años los tipos de interés siguen todavía mínimos y los bancos siguen sin haber adaptado sus modelos de negocio a esta realidad y a los requerimientos regulatorios, lo cual perjudica gravemente sus ingresos y su estabilidad.

De manera recapitulativa, el modelo de Amazon es sin duda un modelo que las empresas del sistema financiero buscan adoptar como el centro de sus estrategias futuras. Mediante el uso de tecnología muy avanzada y el desarrollo del análisis del *Big Data*, los bancos y otros actores financieros, pueden crear productos y servicios con un grado de personalización individual extremadamente alta (adaptados a su perfil de riesgo, nivel de ingresos y capacidad de pago), de manera que captarían a un mayor número de clientes y reducirían sus costes. Sin embargo, esto sigue sin ser suficiente para mejorar la situación actual de los bancos. El modelo de negocio que tiene la mayoría del sector financiero todavía sigue sin adaptarse a las regulaciones crecientes y al extenso periodo de los bajos tipos de interés que no pueden subir debido a los resultados negativos que han tenido los mercados financieros. Además, la competencia creciente de nuevos intermediarios financieros (las conocidas criptomonedas, la financiación de préstamos por medio de vehículos alternativos, etc.) y la desconfianza en el sistema financiero después de la crisis de 2009, hace que redefinir sus modelos de negocio sea más complicado de lo que pensaban.

3.2 Aplicaciones en distintas áreas

En el sector financiero y bancario, existen oportunidades significativas para obtener beneficios mediante la aplicación de las técnicas de minería de datos (*Data Mining*). Más concretamente, cabría destacar tres áreas concretas para la aplicación de las técnicas de *Data Mining*: la seguridad y detección de fraude (mediante la monitorización de registros de transacciones bancarias); la gestión de riesgos y la banca de inversión; aplicaciones en relación con el marketing y la gestión de las relaciones con el cliente como por ejemplo la segmentación de los clientes y las ventas *cross/up*; y, finalmente, en relación con estrategias de expansión y de eficiencia (Hassani, Huang, y Silva, 2018).

- **La seguridad y detección de fraude**

Con el uso generalizado de las tecnologías de la información, entre ellas y sobre todo Internet, la banca online, como ya hemos mencionado, está surgiendo como método y principal medio para el desarrollo de la banca minorista y de la banca comercial. Sin embargo, el crecimiento descontrolado de la información ha provocado, entre otras cosas, un aumento de la complejidad y sofisticación de las actividades fraudulentas, lo cual se considera como un ataque ya que amenaza la seguridad de la entidad y pone en riesgo el futuro de su negocio. (Wei, Li, Cao, Ou, y Chen, 2012). Es más, se afirma que las actividades fraudulentas han supuesto una pérdida anual de más de \$1.744 millones (Amakobe, 2015). En consecuencia, ser capaz de detectar el fraude bancario en Internet se considera un problema importante para todos los bancos (Wei, Li, Cao, Ou, y Chen, 2012). Por ello, se han encontrado diversas posibles soluciones para mejorar los procesos ya existentes de análisis de la información con el objetivo de prevenir y de detectar el fraude.

El análisis de datos se caracteriza por estudiar, con técnicas específicas, el contenido, significado y las características intrínsecas de la información que está a nuestro alcance. Este cuidadoso examen es capaz de identificar las lagunas, los puntos fuertes, las debilidades y los factores de riesgo que pueden llegar a construir amenazas y que finalmente, sugieren líneas de orientación (Bănărescu, 2015). Centrándonos más concretamente en el análisis de datos como sistema implementado para la prevención y

detección de fraude existen más “de 24 tipos de análisis” (Bănărescu, 2015), sin embargo los más relevantes son el análisis operacional y el análisis estratégico (Bănărescu, 2015).

Por un lado, el análisis operacional se caracteriza por estar centrado en el corto plazo ya que implementa y usa información y datos actuales para detectar las actividades presentes que son consideradas como fraudulentas. Este tipo de análisis de datos se centra en “ayudar al administrador antifraude a detectar y combatir actividades ilegales” y, además, mejora las condiciones del lugar de trabajo debido a que intenta evitar actividades manuales y busca reducir el esfuerzo mental (Bănărescu, 2015). Sin embargo, el análisis estratégico implica incorporar un enfoque a nivel macroeconómico al estudio de prevención y detección de fraude. En consecuencia, se estudiará otro tipo de variables, como por ejemplo los riesgos, las tendencias de evolución en los motivos de fraude, la evolución de los mercados financieros, los cambios demográficos o incluso el desempeño y rendimiento de las empresas y de sus actividades. (Bănărescu, 2015). El trabajo se centrará más en describir fenómenos, buscar explicaciones y hacer predicciones, es, por tanto, un análisis hecho para fundamentar las decisiones de la gestión a más alto nivel.

Llega un momento en el que los métodos de investigación clásicos se vuelen ineficientes y técnicos expertos e investigadores deciden aplicar técnicas de *Data Mining* a la búsqueda e identificación de comportamientos inusuales relacionados con el fraude. Dichos investigadores y científicos buscaban, mediante el uso de información de transacciones, encontrar la manera de poder identificar comportamientos sospechosos con la máxima precisión y exactitud posible. Las técnicas de Minería de Datos no son más que técnicas que buscan identificar patrones: todo conjunto de datos, de información tiene una tendencia histórica, y cuando encontramos un elemento que se desvía de ese comportamiento hay que explicar las razones que han causado dicha desviación. Muchas veces estas desviaciones se deben a causas naturales, pero existe la posibilidad de que este comportamiento atípico se deba a una actividad fraudulenta. Por tanto, una de las técnicas más eficientes y más famosas para detectar fraude es buscar el valor atípico de todo conjunto de datos. (Deshpande, Siddiqi, Alam, y Parmar, 2016).

Las técnicas de detección de fraude caen en dos clasificaciones: la Clasificación Supervisada y la Clasificación no Supervisada. En los métodos supervisados, los modelos tienen que diferenciar los comportamientos fraudulentos de los no-fraudulentos, de

manera que cada nueva observación recaiga en una clase o en la otra, proporcionando así un mejor desempeño a la clasificación. Entre los principales problemas de estos métodos encontramos la necesidad de que haya ejemplos anteriores, esto quiere decir que un comportamiento se clasificará como fraudulento solo si en el modelo ya se ha introducido un comportamiento similar anteriormente, si no, no lo detectará. Además, debido al desequilibrio en relación con el tamaño (existen muchos más ejemplos de comportamientos no-fraudulentos que fraudulentos), muchas veces el modelo también puede equivocarse y falsear los resultados. (Deshpande, Siddiqi, Alam, y Parmar, 2016). Los métodos no supervisados simplemente se encargan de buscar los comportamientos inusuales. Son métodos muy utilizados y útiles en aplicaciones en las que no existe ningún conocimiento previo de una clase concreta de observaciones. Para guiar a este modelo se puede construir un modelo de referencia que represente el comportamiento determinado como “normal” para que luego pueda detectar las inconsistencias, “los comportamientos que muestran la mayor desviación de esta norma” (Deshpande, Siddiqi, Alam, y Parmar, 2016).

El autor en (Amakobe, 2015) propone una amplia técnica que es capaz de detectar actividades fraudulentas en tiempo real, y que a la vez es más precisa por medio del uso de la plataforma Hadoop, que a su vez es más rentable y maneja más eficientemente grandes cantidades de datos. Por otro lado, los autores en (Wei, Li, Cao, Ou, y Chen, 2012) proponen una plataforma llamada *i-Alertor* para los bancos australianos más importantes; o, por ejemplo, otros autores proponen un sistema híbrido de minería de datos que predice cuándo se van a producir intrusiones en la red, y que detecta las actividades fraudulentas (Hassani, Huang, y Silva, 2018).

Finalmente, se han propuesto una serie de medidas para implementar acciones de lucha contra el fraude: detectar, responder, investigar y descubrir (Amakobe, 2015). Cuando hablamos de detectar nos referimos a ser capaces, mediante análisis avanzados de la información que disponemos, de predecir si cierta acción es fraudulenta o no. Responder se basa en tomar medidas contra este tipo de acciones en tiempo real; investigar para encontrar respuesta a suposiciones y finalmente, descubrir, gracias al estudio de datos históricos, patrones de fraude y de delitos financieros. (Amakobe, 2015).

- **La gestión de riesgos**

Existen numerosos riesgos en todo tipo de entidad y de organización que ponen en peligro la estabilidad de su negocio y que puede conllevar grandes pérdidas. En las entidades bancarias, existen cinco principales riesgos: el riesgo de crédito, el riesgo del mercado, el riesgo operacional, el riesgo del tipo de interés y el riesgo relacionado con la rentabilidad. Sin embargo, uno de los requisitos más importantes e imprescindibles de todo banco es ser capaz de evaluar propiamente el riesgo inherente del banco debido a que un prestatario pueda o no devolver los pagos del principal y los intereses de la deuda. Para ello, se calcula y se adjunta a cada cliente un *credit score* que es en otras palabras, la capacidad que éste tenga de devolver el préstamo, y en función de éste también se impondrá distintas limitaciones y condicionantes al préstamo.

La forma de medir la calidad crediticia de un demandante de crédito se basa estudios de las relaciones previas con el cliente: mirando su historial, la información disponible sobre su nivel de ingresos, etc. (García Montalvo, 2014). El método más comúnmente usado es el FICO que califica en porcentajes de una nota final distintas variables (el historial de pagos anteriores, el uso del crédito, los tipos de créditos que se han solicitado, etc.). Existe actualmente una tendencia a calificar crediticiamente a los clientes en función de una variable adicional: las conexiones sociales, que considera la reputación, el estatus social y los contactos que se tengan en ciertas redes sociales. Sin embargo, las empresas que miden sus *credit scores* incluyendo esta variable poseen un modelo de negocio diferente que mide el crédito desde otro punto de vista. Un ejemplo es Neo Finance, una empresa que mide “el historial laboral y el número y calidad de las conexiones de LinkedIn con los trabajadores de su empresa para predecir la estabilidad en el empleo y los ingresos futuros” (García Montalvo, 2014).

Las entidades bancarias han decidido no incluir esta variable en sus métodos de valoración de la calidad y capacidad crediticia de sus clientes, pero sí que buscan incorporar los avances del *Big Data* en la predicción de dicha variable. Por ejemplo, BBVA utiliza el pago con tarjeta en TPV para mejorar sus estimaciones.

Resulta que el ámbito de aplicación más importante en el que las técnicas de *Big Data* ven su aplicación en la banca minorista es la valoración de concesión de préstamos a entidades e individuos. Cada decisión de préstamo que toma una entidad bancaria

conlleva una cierta cantidad de riesgo, con lo cual decisiones como la concesión de un préstamo o la ampliación de la línea de crédito puede llegar a ser muy arriesgado si no se conoce bien al cliente (Pascu, 2018). Para solucionar este problema, la minería de datos intenta cuantificar el riesgo asociado con la facilitar de tomar decisiones de crédito, permitiendo y limitando al mismo tiempo las potenciales pérdidas que un banco pueda sufrir (Pascu, 2018).

Si un estudio es capaz formular un modelo en el que se pueda especificar un patrón de comportamiento (que ya haya ocurrido) que lleve al impago de la deuda, entonces, este modelo puede prevenir riesgos futuros cuando se descubran comportamientos semejantes. Las técnicas de *Data Mining* pueden proporcionar a las entidades bancarias numerosas facilidades: permite identificar qué cliente va a retrasarse o incumplir con sus pagos, pero también puede ayudar a analizar los comportamientos y la fiabilidad de los clientes a la vez que venderles tarjetas de crédito, y finalmente, puede también ayudar a ver cierto cliente al que se le ha vendido una tarjeta de crédito, tendrá problemas o retrasos con los pagos que debe. (Subashini y Chitra, 2013).

- **Marketing y gestión de relaciones con el cliente**

En el mundo de la banca, ser capaz de retener a un cliente y de además aumentar la cartera ya existente de clientes es cada vez una tarea más difícil debido a la creciente competencia. Por ello, la única forma de conseguir alcanzar ambos objetivos es mediante el conocimiento de lo que verdaderamente quieren o necesitan los clientes, para así poder llegar a cumplir sus expectativas. (Pascu, 2018). Las técnicas de *Data Mining* aplicadas a la gestión de las relaciones con los clientes permiten analizar la información disponible de un cliente y descubrir indicadores clave que dan directrices a los bancos sobre qué factores han afectado la demanda de sus clientes en el pasado y sus necesidades en el futuro (Pulakkazhy y Balan, 2013). Por tanto, se realiza una clasificación de los clientes en base a varias variables (comportamiento, necesidades, etc.), lo que permite a una entidad enfocar con más precisión sus técnicas comerciales o de marketing. Existen dos modelos principales para clasificar a un cliente: el modelo de clasificación crediticia y el modelo de puntuación conductual (Pulakkazhy y Balan, 2013). Estas clasificaciones

permiten a las entidades bancarias ofrecer distintos servicios financieros a cada cliente dependiendo de su clasificación.

Las técnicas de minería de la información aplicadas a la búsqueda de una mejor gestión de las relaciones con el cliente han mejorado gracias a mejoras en las tecnologías de la información y de las comunicaciones, y al desarrollo de plataformas online, entre otras cosas, y hoy en día permiten a las empresas que las usan anticipar las necesidades de sus clientes. Dicha gestión de las relaciones con los clientes es uno de los principales focos de toda entidad que forma parte del sistema bancario, “siendo, además, parte de la estrategia para construir, administrar y fortalecer las relaciones a largo plazo” (Pascu, 2018). La gestión de las relaciones con los clientes es en sí un negocio a largo plazo que se centra en recoger, analizar y entender información de los clientes para poder tratarlos de forma diferente, y proporcionar un mejor servicio a cada uno de ellos, basado en sus características personales.

La gestión de las relaciones con el cliente (*Customer Relationship Management*) se puede definir en base a cuatro diferentes variables: la identificación de clientes, la captación de clientes, la retención de dichos clientes y el desarrollo del cliente (Pascu, 2018).

Cuando buscamos identificar potenciales clientes ante todo hay que segmentar las clases de potenciales clientes y luego llevar a cabo un análisis de estos. La segmentación se caracteriza por dividir al conjunto de potenciales clientes en grupos más pequeños que compartan características, mientras que el análisis definirá qué clases son más rentables y atractivas para la empresa. La captación de clientes es la etapa posterior a la identificación de los clientes que buscamos captar, y se centra básicamente en implementar todos sus activos en atraer a ese grupo concreto de clientes. Entre los posibles métodos de captación de clientes podemos encontrar la reducción en el precio o la comunicación de las ventajas competitivas que tiene un producto en comparación a otro similar. La fase siguiente es la retención de dichos clientes y es sin duda una de las fases más complicadas de todo el proceso debido a su particular búsqueda de una mayor satisfacción del cliente. La satisfacción que tenga un cliente sobre un producto o un servicio va a determinar la percepción de este sobre el valor que le aporte una empresa. La satisfacción del cliente implica indirectamente su grado de lealtad hacia una empresa,

a mayor satisfacción, mayor lealtad hacia una marca o un producto concreto. Finalmente, el desarrollo del cliente busca aumentar la rentabilidad individual de ese cliente mediante una mejora de las relaciones o mediante la venta cruzada de productos y servicios. (Pascu, 2018).

3.3 Beneficios del *Big Data*

La clave del éxito detrás del *Big Data* no depende de la cantidad de información recogida y almacenada, sino que recae en el uso que una empresa decida hacer de los datos. Los procesos de análisis y de extracción de valor de los datos que tenga una empresa tienen que ser lo más eficientes posible para aumentar las probabilidades de éxito y para conseguir mejores beneficios y resultados. Existen distintos tipos de beneficios que aporta la utilización correcta del *Big Data*:

- **Reducción de los costes**

El *Big Data* proporciona sistemas de gestión para la inteligencia empresarial que permiten mejorar la eficiencia operacional y reducir los costes. Mediante el uso de las tecnologías de datos como Hadoop o las “*cloud-based analytics*” (Balachandran y Prasad, 2017) las empresas pueden ahorrarse costes en relación con el almacenamiento de la información ya que estas técnicas permiten el almacenaje masivo de datos.

Cuando comparamos los sistemas de análisis de *Big Data* con las tecnologías tradicionales, a pesar del alto coste inicial requerido para la instalación de dichas tecnologías más innovadoras, estas proporcionarán a la larga beneficios mayores y, por tanto, el proyecto resulta más rentable. Además, en vez de procesar y almacenar grandes cantidades de datos en los almacenes, las empresas ahora utilizan los *clusters* Hadoop para este propósito y mueven los datos a los almacenes a medida que se necesitan para aplicaciones analíticas del proceso de producción. (Balachandran y Prasad, 2017).

- **Mejora en la toma de decisiones**

El procesamiento cada vez más rápido de las tecnologías Hadoop hacen que las empresas puedan procesar y analizar instantáneamente la información que ya poseen para realizar predicciones del futuro. Además, también buscan ser capaces de encontrar nuevas fuentes de información que complementen sus conclusiones y recomendaciones y que

facilite, por tanto, el proceso de la toma de decisiones haciéndolo mucho más efectivo y argumentado.

Cuando tenemos datos en tiempo real sobre clientes, promociones, tendencias, gustos, ofertas competitivas, comportamientos de la competencia, etc., las empresas pueden tomar decisiones más informadas y más rápidas en el presente y establecer una estrategia para prepararse para el futuro.

- **Oferta de productos y servicios más personalizados**

El análisis de la información recogida y recopilada por una empresa otorga la capacidad de identificar patrones de sus clientes. Entender qué productos o servicios funcionan mejor o peor y el por qué detrás de cada elección de compra, permite estimar factores como la satisfacción y las necesidades de los clientes. Mediante el análisis histórico de los comportamientos de compra y de las reacciones de los clientes a nuevos lanzamientos de producto permite que las empresas sepan exactamente qué ofrecer para tener éxito. Además, el análisis del comportamiento de los mercados (cambios en la oferta y demanda de productos) en tiempo real también ayuda a crear un contenido de marketing más orientado al cliente.

Las empresas terminan, por tanto, ofreciendo productos y servicios que no solo satisfagan en un mayor grado las necesidades de sus clientes, sino que muchas veces les abren los ojos mediante posibilidades de nuevos productos que no sabían que necesitaban. Entender las necesidades de un consumidor, estudiar sus comportamientos y preferencias puede ayudar a crear servicios y productos mucho más personalizados.

En el caso de los bancos, es común la pérdida de clientes debido a la poca interacción personal con ellos, es decir, debido a la falta de conexión emocional entre ambas partes. Los gestores de relaciones o las partes interesadas en captar clientes o vender productos y servicios financieros pueden identificar, mediante el análisis histórico de los datos de sus clientes, patrones de inversión, antecedentes financieros y personales, motivaciones personales para invertir, de manera que ofrezcan soluciones de inversión (cuentas, seguros, préstamos, etc.) personalizadas. De este modo, los bancos ofrecerán verdaderamente los servicios requeridos por y para sus clientes. (Kathuria, 2016).

Finalmente, los bancos saben que el coste de adquirir nuevos clientes es mayor que el coste de retener los clientes ya existentes. La recogida de información en tiempo real y

el análisis histórico de esta permite a los bancos tomar medidas antes de tiempo para evitar potenciales pérdidas.

- **Detección de fraude y cumplimiento de la regulación**

El análisis de *Big Data* ayuda a que las empresas detecten automáticamente los intentos de fraude para hackear los sistemas de su organización. Analizando diferentes amenazas internas, las organizaciones pueden mantener su información confidencial segura almacenándola, siguiendo los requisitos reglamentarios. Se trata, además, del principal problema al que se enfrenta el sector bancario. (Kathuria, 2016).

3.4 Inconvenientes y peligros del *Big Data*

La utilización y adaptación de las técnicas relacionadas con *Big Data* al modelo de negocio de una empresa pueden suponer una gran ventaja competitiva que aumente la satisfacción de sus clientes y, en consecuencia, aumente sus beneficios y rentabilidad a largo plazo. Sin embargo, la utilización de dichos procedimientos y sistemas también consta de algunos peligros.

- **Almacenamiento**

En un primer momento, el principal riesgo de la implementación de técnicas relacionadas con el *Big Data* tiene que ver con el almacenamiento de la información. Ser capaz de almacenar grandes cantidades y volúmenes de datos es crucial para que una empresa pueda funcionar y ello requiere que dicha empresa tenga una infraestructura de hardware y software amplia y compleja.

Con el crecimiento del volumen de información disponible al alcance de todo tipo de empresa y de negocio, los grandes actores buscan aumentar sus capacidades de almacenamiento y procesamiento en la nube para poder llegar a ser aún más competitivos. (Balachandran y Prasad, 2017).

- **Posibilidad de fracaso**

Las tecnologías de análisis de *Big Data* proporcionan herramientas muy útiles para las empresas ya que son capaces de aportar información de alto valor mediante el reconocimiento de tendencias, comportamientos clave, etc. Sin embargo, el éxito de su

aplicación e integración a la empresa no está garantizado y es posible que termine resultando en un ejercicio costoso sin ninguna salida.

Cabría destacar que la creación de valor no se obtiene por medio de una acumulación de una cantidad mayor de información, ya que la información redundante y repetitiva no permite solucionar ningún problema. “Los datos por sí mismos no proporcionan una ventaja competitiva a no ser que el análisis sea el adecuado, por lo que es muy importante contar con un equipo profesional de *analytics* que pueda extraer conclusiones apropiadas a partir de los datos” (García Montalvo, 2014). Pero también resulta importante el conocimiento del ámbito de aplicación de dichas técnicas. Es decir, si una empresa no entiende de qué manera el *Big Data* puede aportar beneficios a su negocio, si no entiende cómo debe de utilizarlo, lo implementará de la forma errónea y eso le llevará a tomar decisiones inapropiadas y a tener pérdidas.

- **Privacidad y Seguridad**

El avance de las tecnologías de la comunicación ha provocado que compartamos y se almacene información personal sobre cada individuo casi constantemente, a través de la red móvil, de compras en el supermercado, de visitas al médico, de la compraventa de servicios financieros, de las redes sociales, de los datos que poseen las empresas sobre sus empleados, etc. Esto ha provocado que la privacidad se haya convertido en un tema de primordial importancia para todos los afectados, especialmente para los clientes.

Cada vez hay más clientes que rechazan compartir su información por temor al uso inapropiado de los datos personales. Y es que la rapidez con la que se producen datos nuevos y las probabilidades de que se generen situaciones difíciles y paradójicas hace que las leyes de protección de datos siempre vayan un paso por detrás de la realidad. Además, también existen la paradoja y complicada relación entre las opiniones sobre la privacidad personal como derecho humano fundamental. Un claro ejemplo es la ley de protección de datos del Reino Unido, que protege la información a nivel nacional pero no es aplicable a ningún tipo de información personal que se haya almacenado fuera del territorio nacional, y, sin embargo, sus habitantes tratan diariamente con empresas con exposición y escala global. (Oguntimilehin y Ademola, 2014).

La mayoría de las empresas hoy en día obligan a sus clientes o usuarios completar una serie de cláusulas de consentimiento, no solo para la recogida y almacenamiento de

datos sino también para su uso y su análisis. Sin embargo, la mayoría de los clientes no se leen dichas cláusulas y, por tanto, no saben lo que están firmando y aún menos a lo que están accediendo. Finalmente, otro de los elementos situados en el centro del debate actual es el valor de la información, ya que todavía las empresas no pagan a sus clientes por el uso de su información y de sus datos personales. (García Montalvo, 2014).

Para sacarle la mayor rentabilidad y partido a los sistemas de análisis de datos, las organizaciones deberían establecer políticas de seguridad que sean auto configurables. Estas políticas deben aprovechar las relaciones de confianza existentes entre entidades, promover el intercambio de datos y de recursos entre ellas y asegurarse que el análisis de la información se optimice en vez de que se limite debido a dichas políticas. Y en el caso de que se produjeran ataques cibernéticos que puedan poner en peligro la seguridad de la información de todos los clientes, las empresas deberían involucrarse en la compra de hardware de seguridad para rastrear el comportamiento inusual de actores en los servidores e instalar aplicaciones de seguridad. (Balachandran y Prasad, 2017).

- **Ausencia de calidad de la información**

La precisión y disponibilidad instantánea de la información es crucial para garantizar una efectiva toma de decisiones. Los procesos de *Big Data* generan valor únicamente cuando ha habido un proceso que garantice la calidad de la información y de los datos que se van a procesar.

En caso de que se utilicen datos erróneos, puede afectar negativamente no solo a los clientes, sino que también puede llegar a incumplir alguna regulación. En relación con las consecuencias negativas para los clientes se podría resaltar el ejemplo de las empresas de generación de *credit scores* a partir de información sacada de Internet. “Un estudio de 2013 de la Fed (*Federal Trade Commission*) de Estados Unidos señalaba que el 20 por ciento de los informes crediticios de este tipo de compañías contienen errores, y un 5 por ciento de estos errores resultaron en una rebaja del *credit score* que impidió a los clientes conseguir un crédito o les supuso pagar un tipo superior” (García Montalvo, 2014).

4. Banca en el futuro

Las extremadamente avanzadas tecnologías de los sistemas informáticos y digitales han demostrado tener una presencia notable en el actual modus-operandi del sistema bancario. Las tecnologías líderes incluyen *Big Data Analytics*, *Cloud Computing*, la Inteligencia Artificial (AI), *Machine Learning* (ML) y el Internet de las Cosas (*Internet of Things*, IoT) (Ravi y Kamaruddin, 2017).

La industria financiera está adoptando ya nuevas tecnologías y estableciendo precedentes al respecto. Gracias a la recogida masiva de información a la que tienen acceso los bancos por medio de sus servicios tradicionales, las instituciones bancarias han evolucionado no solo con el objetivo de analizar históricamente la información de la que disponen, sino también ser capaz de manejar datos en tiempo real. Lo que buscan es entonces aprovechar sus fuentes de datos históricas y en vivo para integrarlo en sus sistemas operacionales y analíticos y ser capaz de entender y obtener directrices de la creación de valor de su negocio. Un ejemplo concreto es el sistema de negociación por medio de algoritmos, que usa tanto datos históricos del comportamiento del cliente (en los que establece patrones de comportamiento) como datos en tiempo real de la negociación. O incluso el uso de las tarjetas de crédito, que ya hemos explicado en apartados anteriores, en el que se busca detectar los comportamientos fraudulentos mediante un estudio de las actividades históricas del cliente. (Ravi y Kamaruddin, 2017).

Las nuevas tendencias están ahora enfocadas, entre otras cosas, hacia el *Internet of Things* (IoT), en el que se busca generar datos e información en tiempo real desde varios dispositivos que luego se retransmite hacia un sistema central para garantizar un análisis verídico; hacia el *Blockchain* y hacia la Inteligencia Artificial.

4.1 *The Internet of Things* (IoT)

El increíble aumento del número de personas, y, por tanto, de dispositivos que se conectan constantemente entre sí, provocará que el Internet de los Objetos sea la principal fuente de información que todas las empresas tendrán para conocer mejor a sus clientes. Y los bancos no son menos.

El Internet de los Objetos o *Internet of Things* es una tecnología potencialmente disruptiva que busca reconsiderar los procesos y los modelos tradicionales de negocio en una amplia gama de industrias y de sectores. La Unión Internacional de Telecomunicaciones define el Internet de los Objetos como “una infraestructura global de la sociedad de la información, que permite ofrecer servicios avanzados mediante la interconexión de objetos (físicos y virtuales) gracias a la interoperabilidad de tecnologías de la información y comunicación presentes y futuras” (Unión Internacional de Telecomunicaciones, 2012).

La estructura y funcionamiento de esta tecnología se hace a través de diferentes tipos de sensores que están integrados en los dispositivos conectados a Internet que se encargan de recoger y recopilar datos que luego compartirán a través de Internet con personas, aplicaciones, empresas, etc. Los dispositivos IoT (*Internet of Things*) pueden ser cualquier tipo de dispositivo digital, desde una Tablet, un *smartphone*, un sensor transportable y que nos podamos poner o un sector incluso industrial. Una vez recogida la información se retransmite a través de cualquier ruta de red, cableada o inalámbrica. La capacidad de analizar los datos recogidos para poder buscar nuevas oportunidades de negocio y aplicar dichas ideas de negocio tiene el potencial de mejorar el negocio del sistema bancario (Ravi y Kamaruddin, 2017).

Las empresas líderes y predominantes en el desarrollo de estos dispositivos son básicamente Apple, Google, Amazon y Samsung que han desarrollado ya dispositivos ponibles y los primeros dispositivos con voz (son dispositivos donde la interfaz principal es la voz y sirve tanto como para la entrada y salida de información) creados para ser asistentes personales. Sin embargo, el caso más brillante de aplicación de esta tecnología innovadora es la forma en la que Uber ha integrado su capacidad de análisis geoespacial con su capacidad para analizar los precios de oferta y demanda en tiempo real y con su sistema de pago unificado para ofrecer un servicio único y superior de transporte alternativo (Ravi y Kamaruddin, 2017).

4.2 *Internet of Things* (IoT) en la Banca

¿Cuál es el impacto que ha tenido la aparición del Internet de los Objetos en la Banca? Una encuesta global afirma por un lado que el 65% de los ejecutivos bancarios realiza el

seguimiento de sus clientes mediante el uso de aplicaciones móviles; el casi 32% de las instituciones bancarias utilizan el Internet de los Objetos para observar las localizaciones y el funcionamiento de sucursales bancarias, entre otras cosas; y que el 21% de los bancos utiliza sensores para recopilar datos e información del rendimiento de sus productos (Ravi y Kamaruddin, 2017). La realidad detrás de estos resultados es que las instituciones financieras buscan proporcionar y dar máxima prioridad a sus clientes y al control de sus productos y servicios debido al creciente fraude financiero. Ahora más que nunca, los bancos han puesto regulaciones en relación a la verificación de la identidad de todos sus clientes, ya que el robo de la identidad también es una de las mayores fuentes de fraude y de brecha de los sistemas informáticos. (Ravi y Kamaruddin, 2017).

El Mckinsey Global Institute (Manyika, y otros, 2013) estimaba en 2013 que el uso del Internet de los Objetos tendrá un potencial impacto económico anual (en términos brutos) de entre US\$ 2.700 millones y US\$ 6.2 billones para 2025, proveniente de una reducción de los costes y un aumento de la productividad como resultado de la utilización y aplicación de estas tecnologías. Un ejemplo del aumento de la productividad se ha dado en la empresa UPS, operador logístico, que declaraba una mejora de su productividad gracias a una más eficiente supervisión de sus camiones de reparto y sus equipos de entrega. Decía que un conductor normalmente podía entregar 90 paquetes al día, y que, gracias a la optimización de sus sistemas, ahora puede entregar y distribuir hasta 120 paquetes al día (Goldman Sachs, 2014).

Por ello, podríamos decir que los bancos tendrán mejoras en su productividad y en los costes debidos a un incremento en la eficiencia de las operaciones gracias a la implementación de dispositivos capaces de identificar a un cliente a partir de datos biométricos. El mecanismo es parecido a lo que ocurre en los edificios o en los hogares en los que cuando se instala una mayor automatización y conectividad, la eficiencia energética mejorará. No obstante, los bancos también buscarán tomar medidas que proporcionen un aumento en el flujo de ingresos de su negocio.

A pesar del alto coste que acarrea captar a nuevos clientes, es y será una de las estrategias más buscadas por las entidades bancarias. Los clientes que ahora buscarán los bancos estarán centrados principalmente en captar a *millennials* o a personas pertenecientes a la generación Z gracias a un mayor conocimiento y comprensión de sus

características. Los bancos tienen una importante ventaja competitiva en relación con la captación de nuevos clientes ya que disponen de una vasta cantidad de datos y de información sobre las distintas pautas de consumo y comportamientos de los usuarios. (Martin, y otros, 2017). Sin embargo, el problema es ser capaz de mantener a esos nuevos clientes. Esta dificultad proviene de las tendencias actuales que hacen que las personas puedan tener un mayor control y empoderamiento de su información personal. (Martin, y otros, 2017).

Resulta importante averiguar qué medidas pueden tomar las entidades bancarias para captar a las nuevas generaciones de clientes y para asegurar su lealtad en el medio y largo plazo. En primer lugar, hay que entender que cada generación de clientes es diferente de la anterior y por ello, el acercamiento a ellos debe ser distinto. Los bancos que tendrán éxito serán aquellos que consigan transmitir la mejor experiencia a sus clientes. Por otro lado, los *millennials* han nacido prácticamente con un smartphone en sus manos. En consecuencia, la industria financiera y los bancos tienen la oportunidad de jugar un papel clave en la interacción digital de los usuarios. Un ejemplo de adaptación que ha conseguido captar clientes son las aplicaciones de pago, transferencias por móvil. Se dice que en el futuro todo dispositivo tecnológico será capaz de realizar pagos online (por ejemplo, coches que paguen la gasolina). Finalmente, los bancos tienen que jugar un papel primordial en la captura de la información de sus clientes en tiempo real a través del Internet de los Objetos ya que eso les dará una ventaja comparativa esencial para asegurar su futuro. (Martin, y otros, 2017).

En definitiva, para los bancos no solo supone un nuevo medio de captación de clientes, pero también nuevas oportunidades para redefinirse. Esto se transmite en que los clientes ya no tendrán que acudir físicamente al banco o a la sucursal bancaria para contratar un servicio o producto que deseen, sino que los bancos podrán ofrecer una experiencia mucho más completa y más personalizada ya que tendrán un mayor control y seguimiento de las finanzas personales de sus clientes. Es decir, conocerán verdaderamente a sus clientes.

Sin embargo, el Internet de los Objetos todavía tiene ciertas dificultades que tiene que superar. En primer lugar, esta tecnología carece de normas de seguridad eficaces y no recibe ningún tipo de parche de seguridad ni actualización (Martin, y otros, 2017).

Además, también se plantea la crítica de si la recogida de datos y de información puede dar lugar a la discriminación de precios. Debido a que las empresas tienen información muy personal sobre sus clientes y sobre sus comportamientos de compra, podrían entonces excluir a determinados perfiles de determinados segmentos del mercado y elevar el precio de los productos a clientes con mayor poder adquisitivo. (Martín, y otros, 2017).

- ***Internet of Things* y la gestión de relaciones con el cliente**

El Internet de los Objetos se caracteriza por recolectar información automáticamente mediante objetos físicos (sensores) que están conectados a una red global y que transfieren los datos a un repositorio central. Los datos se almacenan y se analiza su significado para poder utilizarlo en el proceso de toma de decisiones, permitiendo así que se tomen decisiones más precisas, adecuadas y argumentadas. Además, existen otros usos, que veremos a continuación, que pueden constituir potenciales fuentes de beneficios para las organizaciones mediante el uso del *Internet of Things*.

En un primer momento, el Internet de los Objetos oferta productos y servicios más personalizados gracias a un estudio de los patrones de compra, preferencias, de la ubicación geográfica y del contexto actual de cada cliente. El objetivo último de esta técnica es ser capaz de mantener al cliente comprometido con la organización y satisfecho mediante el nivel más alto de personalización posible. Pero, además, permite a las organizaciones fijar precios en función de la demanda en tiempo real que tengan sus productos o servicios. Ciertos negocios son extremadamente sensibles al precio, y por ello, el precio adecuado en el momento adecuado se convierte en la clave para garantizar su éxito. Esta técnica tiene la capacidad de hacer promociones “*on-the-fly*” usando flujos de datos que se produzcan en vivo. Ya que el Internet de los Objetos permite recoger información en vivo de cada persona, ahora las empresas pueden ofrecer precios diferenciados para cada cliente en función de sus necesidades. Los sensores recogen en vivo enormes cantidades de flujos de datos y las organizaciones deben ser capaces de saber cómo utilizar esa información para tener una ventaja comparativa sobre sus competidores. (Yerpude y Singhal, 2018).

4.3 Blockchain

El *Blockchain* ha surgido como la próxima generación del sistema y canal de pago sin la necesidad de que un tercero participe o se involucre en el proceso para asegurar su correcto funcionamiento. *Blockchain* es una estructura de datos que permite crear un libro de transacciones digital y compartirlo entre una red de ordenadores. Utiliza la criptografía para permitir que cada participante de la red pueda manipular el libro de forma segura y sin que haya un ente central que controle todo. Se trata de una base de datos de transacciones seguras compartida por todas las partes de la red distribuida, y registra y almacena cada transacción que se produce en la red creando así un historial de transacciones irrevocable y auditable (IBM, 2016). Esto quiere decir que cuando un bloque de datos se registra en el libro o en el sistema, es extremadamente difícil de cambiar o de eliminar.

Esta tecnología ha surgido con el objetivo de que las empresas puedan realizar y verificar transacciones financieras en una red distribuida de forma instantánea y sin que haya una autoridad central. Se trata de un modelo innovador ya que hasta el momento las transacciones financieras o bancarias siempre dependían de un intermediario para que se pudieran autorizar los pagos. Con lo cual, lo que se ha buscado con el *Blockchain* es ser capaz de crear una red distribuida de ordenadores que actúan como intermediarios y que, por tanto, no necesitan de éste. (Fintech Network, 2019).

Uno de los aspectos interesantes de la tecnología Blockchain es el concepto de contratos inteligentes en donde las reglas implícitas están ya incrustadas en la cadena de bloques y se ejecutan con la transacción. Esta tecnología común permite que se reduzcan las fricciones creadas en las redes financieras cuando diferentes intermediarios utilizan distintas infraestructuras tecnológicas. Y finalmente, permite también reducir la necesidad de los intermediarios para validar las transacciones financieras (IBM, 2016).

Esta tecnología ofrece a las empresas financieras la oportunidad de reforzar sus estructuras y de acelerar los acuerdos y los procesos. Entre otras cosas permitirá transferir fondos de forma segura sin necesidad de que haya una confirmación y permitirá manejar pagos transfronterizos, operaciones de capital e incluso liquidación de obligaciones. (Ravi y Kamaruddin, 2017).

- ***Blockchain* y la detección de fraude**

Como ya hemos explicado en apartados anteriores, uno de los principales desafíos a los que se enfrenta la industria bancaria hoy en día es al crecimiento del fraude y de los ciber-ataques. Tradicionalmente, los libros de contabilidad y se creaban y almacenaban en una base de datos centralizada lo cual facilitaba la entrada a *hackers* y que se produjesen ciber-ataques. A pesar de la mejora de los sistemas de seguridad implantados para prevenir estos comportamientos, los *hackers* han seguido evolucionando sus técnicas de manera a sobrepasar cualquier obstáculo. La tecnología del *Blockchain* por su parte, debido a que es un sistema descentralizado, abre menos puertas a que se produzcan comportamientos fraudulentos o ciber-ataques. En definitiva, la implementación de las tecnologías de *Blockchain*, permitirían no solo ejecutar pagos en tiempo real, sino que aportarían una mayor transparencia lo cual mejoraría los sistemas de análisis y prevención de fraude también en tiempo real (Fintech Network, 2019).

Resulta interesante determinar la particularidad presente en la tecnología Blockchain que permite almacenar todo tipo de información y mediante la cual, todos los usuarios presentes en la red pueden acceder a esa información de manera igualitaria. Esto quiere decir que la información presente en la red distribuida de ordenadores es revisada constantemente y en tiempo real por cada participante de manera independiente. El libro de contabilidad que se produce sin la necesidad de un intermediario registra entonces todas las transacciones históricas de cada cliente, lo que permite que los comportamientos sospechosos se detecten y se hagan visibles automáticamente.

4.4 La Inteligencia Artificial

La Inteligencia Artificial se refiere al desarrollo de máquinas o de sistemas que puedan realizar tareas complejas que tradicionalmente se pensaba que requerían de inteligencia humana para realizarse. Más concretamente, es un sistema informático que puede sentir, comprender, actuar y aprender. Se trata de un sistema que puede percibir lo que le rodea, y puede analizar y comprender la información que recibe a la vez que tomar medidas basadas en dicha comprensión. Este tipo de tecnología va a permitir ampliar las capacidades de los humanos y de las máquinas mucho más allá de lo que harían por separado. (Accenture, 2018). La Inteligencia Artificial proporcionará herramientas a los

bancos para mejorar su eficiencia mediante una reducción de los costes, una mejor mitigación del riesgo y mediante un aumento de los ingresos gracias a la aplicación de nuevas tecnologías en el análisis y en los *bots*, entre otras. (Latimore, 2018).

El progreso del análisis que hemos mencionado en el apartado anterior se explica por medio de tres factores: el poder de procesamiento, la disponibilidad de información y la mejora de los algoritmos existentes. El poder de procesamiento requerido para satisfacer ciertas demandas es simplemente demasiado grande: el crecimiento desmesurado de la información hace que sea actualmente imposible que un humano pueda recuperar y analizar los datos necesarios para dar respuesta a una transacción. Sin embargo, los bancos ya poseen una ventaja comparativa sobre otras industrias ya que de por sí ya procesan enormes cantidades de datos sobre sus clientes. Si éstos son capaces de combinar los datos con información sobre los comportamientos y con información proveniente de fuentes externas, entonces serán capaces de desarrollar una relación mucho más personal con sus clientes. Los datos son el componente más importante para la construcción del *machine learning*, y el análisis correcto es el que utiliza enormes cantidades de datos. Finalmente, la mejora automática de los algoritmos resulta indispensable. Esto quiere decir que se necesita que cada vez que ocurra un error, un programa que utilice la tecnología de *machine learning* realice los ajustes necesarios. (Latimore, 2018).

Por otro lado, el ejemplo más visible hoy en día de la Inteligencia Artificial es el conocido “*Chatbot*”. Es un servicio con el que se puede comunicar a través de mensajes de texto. El *Chatbot* recibe el mensaje, entiende lo que se está tratando de decir y responde con un mensaje coherente o directamente realiza la tarea que el cliente desea (Latimore, 2018). La mayoría de las veces se presentan en forma de agente virtual que puede tener un nombre o algún tipo de personalidad, es como si hablásemos de un Avatar. Se accede a ellos mediante las propiedades digitales del propio banco y actuarán como intermediario con el banco o ejercerán funciones tales como las de un *callcenter*. (Latimore, 2018).

Sus ventajas competitivas son que permitirán a toda empresa que los implemente reducir el tiempo en la realización de cualquier actividad, también reducirán los costes laborales y aumentarán la eficiencia. En relación con la industria financiera, los *chatbots* tienen distintas aplicaciones que permiten crear valor: se puede emplear al *chatbot* como

un asistente de Recursos Humanos, o como un asistente de inteligencia de los mercados, o incluso como embajador de marca para el negocio. Según un estudio de investigación, el 80% de las instituciones financieras a nivel global considerarán la implementación de los *chatbots* para mejorar la productividad de sus respectivos negocios (Ravi y Kamaruddin, 2017).

4.5 Impacto de la Inteligencia Artificial en la Banca

Como hemos explicado anteriormente, la Inteligencia Artificial tiene el potencial de mejorar todas las áreas y departamentos de un banco (desde el *Front Office* que trata e interacciona directamente con el cliente, como el *Middle Office* y el *Back Office*). Estudiaremos los beneficios y riesgos que aporta la integración y aplicación de las tecnologías relacionadas con la Inteligencia Artificial en las distintas áreas del ámbito bancario.

- ***Front Office***

Cuando hablamos de *Front Office* nos referimos a todas las estructuras de una organización que gestionan las relaciones con el cliente. Ciertas partes de las relaciones con el cliente pueden ser abordadas por la Inteligencia Artificial directamente, como con *chatbots* o agentes virtuales, o con estructuras y sistemas que permitan a los empleados realizar un trabajo mejor (ofrece una estructura que hace que hagan su trabajo más rápido, más precisamente y más eficientemente). La interacción con el cliente puede unilateral (la empresa lanza numerosas ofertas a los clientes, pero no se espera respuesta por su parte) o una relación bilateral (continuo intercambio de información por las dos partes). Entre los principales beneficios para los clientes encontramos una mejora en las ofertas de productos y servicios financieros, pero también les permite ahorrar tiempo significativamente. Sin embargo, si la implementación resulta en defectuosa puede terminar costándole dinero al banco o también puede acarrear un riesgo reputacional (empeora la percepción que tiene el cliente hacia la entidad bancaria) de tal manera que el cliente no querrá volver a tener contacto con los sistemas de Inteligencia Artificial en el futuro. (Latimore, 2018).

- ***Middle Office***

Por otro lado, la parte del banco enmarcada en el conocido “*Middle Office*” se encuentran todos los empleados que ofrecen apoyo indirectamente a los clientes o que también realizan actividades administrativas correspondientes en parte al *Back Office* o de cumplimiento. Las principales actividades que realizan son: la generación de reportes, decisiones de crédito y de riesgo y gestionan la parte relacionada con el cumplimiento de las normativas legales por parte de la empresa. La Inteligencia Artificial permitirá a los trabajadores del *Middle Office* hacer más trabajo a un menor coste, es decir, buscará potenciar la eficiencia de su trabajo. Sin embargo, siempre tendrá que tener a una persona detrás de este trabajo para revisar y asegurar el correcto cumplimiento de las tareas. (Latimore, 2018).

- ***Back Office***

El *Back Office* son el conjunto de actividades de apoyo al negocio, las que están detrás y que dan soporte a la gestión empresarial. Funciones típicas de esta parte son el procesamiento y la conciliación de información, y la Inteligencia Artificial puede permitir encontrar datos discordantes, es decir, podrá ayudar a determinar anomalías. Los clientes no verán beneficios muy altos en esta parte del banco, sin embargo, el principal riesgo recae en que el banco confié demasiado en este sistema además del alto coste de implementación de estos sistemas. (Latimore, 2018).

- ***Inteligencia Artificial y la gestión de riesgos***

El riesgo de crédito es el riesgo que se materializa en la pérdida económica procedente del incumplimiento de una contraparte de sus obligaciones contractuales. Se trata del riesgo de impago de una transacción y para calcularlo, tradicionalmente, las entidades bancarias utilizaban regresiones lineales que realizaban modelos en los que basaban sus predicciones. Sin embargo, la aparición de la Inteligencia Artificial y de las técnicas de aprendizaje automático, han suscitado el interés de todo el sector bancario. La realidad es que las técnicas clásicas pueden mejorarse gracias a la implementación de estas novedosas tecnologías debido a su capacidad de comprensión de los datos no

estructurados. Se trata de la dificultad particular de materializar el riesgo de crédito lo que ha hecho que las puertas de la Inteligencia Artificial se hayan abierto. Un claro ejemplo es en el mercado de *credit default swap* (CDS) donde es muy difícil estimar tanto la probabilidad de incumplimiento del pago como estimar el coste del incumplimiento, debido a la incertidumbre existente en numerosos eventos. (Aziz y Dowling, 2018).

5. Conclusiones

La revolución del *Big Data* ha venido con grandes oportunidades y desafíos. Cuando hablamos de *Big Data* nos referimos a las técnicas que empresas y otros actores utilizan para analizar y procesar grandes volúmenes de datos (estructurados, no estructurados y semiestructurados) que se producen todos los días para aplicarlo a su toma de decisiones y a sus estrategias de negocio. La cantidad de datos que producimos diariamente simplemente escapa a la mente humana y cabe resaltar que el noventa por ciento de todos los datos que tenemos hoy en día en el mundo se han generado en los dos últimos años. De media, Google procesa 40.000 búsquedas por segundo, lo que son 3.5 billones de búsquedas al día y Facebook tiene una actividad que hace que 1.5 billones de personas se conecten al día a su plataforma. Cada minuto se reproducen más de cuatro millones de videos en YouTube, se envían más de 400.000 tweets y se publican más de 46.000 fotos en Instagram, cada minuto (Marr, 2015).

Sin embargo, resulta relevante explicar los componentes inherentes al análisis del *Big Data*, es decir, las características que permiten su existencia y su consecución. El primer componente que forma la ciencia del *Big Data* es la obtención de la información y las distintas fuentes usadas en la actualidad. Existen distintos métodos de recolección de datos como por ejemplo las bases de datos tradicionales, encuestas u observaciones, no obstante, como se puede deducir del apartado anterior, la mayoría de los datos se recogen gracias al uso de las redes sociales. En segundo lugar, uno de los requisitos más problemáticos y a la vez relevantes para el análisis del *Big Data* es el almacenamiento de la información. Debido al crecimiento exponencial de producción de datos y de información, se necesitan de plataformas estables y capaces de soportar grandes volúmenes de datos. Entre los métodos de almacenaje de la información podemos asimismo distinguir entre métodos tradicionales (bases de datos relacionales) y más innovadores que permiten adicionalmente almacenar los conocidos datos no estructurados y semiestructurados (*Datawarehouse* y NoSQL).

Por otro lado, para cumplir el objetivo de sacar información de valor de los datos, es importante llevar a cabo la técnica adecuada de análisis y procesamiento de la información y de un conjunto de datos. En esta fase del proceso también encontramos técnicas tradicionales y técnicas más actuales. Entre las tradicionales encontramos la

Minería de datos que consiste en la limpieza, la integración, la selección, la transformación, la extracción y la minería de datos para que finalmente se evalúe y se presente el conocimiento. Sin embargo, el análisis del *Big Data* difiere del tradicional en que los métodos utilizados para analizar datos tradicionalmente son incompatibles con el ritmo de crecimiento actual de la información. Por ello, se creó el procesador de infraestructura llamado Hadoop que tiene numerosas aplicaciones, entre ellas en el conocido buscador Yahoo!

Los nuevos beneficios que aporta el análisis del *Big Data* son sin duda la rapidez y la eficiencia. Mientras que hace unos años una empresa habría recolectado información y habría realizado un análisis sobre la información que podría ser usada para decisiones futuras, hoy en día esa misma empresa puede identificar por medio de los datos, nuevas ideas y conclusiones que le permitan tomar decisiones inmediatas. La capacidad de trabajar más rápido y permanecer ágil da a las organizaciones una ventaja comparativa que antes no tenían.

El análisis del *Big Data* ayuda a las organizaciones a saber cómo aprovechar sus datos y utilizarlos para identificar nuevas oportunidades. Esto, a su vez, conduce a movimientos empresariales más inteligentes, operaciones más eficientes, mejores beneficios y clientes más satisfechos. Esta tecnología aporta valor de distintas formas:

1. **Permite reducir los costes.** Las tecnologías relacionadas con el *Big Data* como por ejemplo Hadoop o *Cloud-Based Analytics*, aportan ventajas significativas en relación con los costes de almacenar grandes cantidades de datos, y, además, pueden llegar a identificar distintas formas de hacer negocios de manera más eficiente.
2. **Una mejor y más rápida toma de decisiones.** Con la velocidad que aporta Hadoop y la capacidad de analizar distintas fuentes de datos, las empresas son capaces de analizar información inmediatamente y tomar decisiones basadas en lo que han aprendido.
3. **Nuevos productos y servicios.** Gracias a la nueva capacidad de medir la satisfacción de sus clientes a través del análisis de los datos, las empresas tienen ahora el poder de dar al cliente lo que necesiten para satisfacer sus necesidades.

El mundo financiero y el sistema bancario son asimismo participantes en esta revolución tecnológica. Gracias al desarrollo de las tecnologías y de las comunicaciones, en la actualidad los clientes no tienen que acudir físicamente a una sucursal bancaria para realizar un trámite o ejecutar una transacción financiera. Ahora los clientes pueden manejar autónomamente sus propias transacciones y operaciones gracias al uso de la plataforma *e-commerce*. A su vez, los bancos ahora tienen acceso a una cantidad de información diferente: deben capturar y analizar las acciones realizadas por sus clientes con el fin de entender sus comportamientos, sus necesidades y sus preferencias. Por tanto, una mayor facilidad de acceso y una simplificación de las operaciones ha hecho que el volumen de actividad de las entidades bancarias aumente significativamente. Sin embargo, las tecnologías existentes no eran capaces de procesar las nuevas fuentes de información y por ello, los bancos han tenido que adaptarse e incorporar las técnicas de análisis de *Big Data* para asegurarse una clara ventaja comparativa frente a sus competidores.

Una vez se integran las técnicas de *Data Mining*, en las entidades bancarias tendrían tres impactos clave: reforzar la seguridad y a la detección de fraude; dar soporte a la gestión de riesgos; y, mejorar sus estrategias de marketing y sus relaciones con los clientes. En cuanto a la detección de fraude, científicos y expertos buscaban, mediante la información disponible sobre transacciones, encontrar la manera de poder identificar comportamientos sospechosos con la máxima precisión y exactitud posible en tiempo real. En relación con la gestión de riesgos, las técnicas de Minería de Datos han desarrollado y formulado modelos basándose en patrones de comportamientos fraudulentos a lo largo del tiempo, que dan la posibilidad a los bancos de identificar comportamientos y otras actividades sospechosas en tiempo real, de manera a prevenir riesgos futuros. Finalmente, debido al surgimiento de nuevos competidores, ser capaz de captar clientes y de retenerlos se ha hecho cada vez más complicado en el sector bancario. Por ello, los objetivos de los bancos se han reformulado y se centran actualmente en adquirir conocimiento sobre las necesidades de sus clientes para ser capaz de anticiparlas y crear productos o servicios que las satisfagan. Las técnicas de *Data Mining* han formulado asimismo dos modelos de manera a clasificar a sus clientes: el modelo de clasificación crediticia y el modelo de puntuación conductual para así ofrecer unos productos y servicios financieros más personalizados y a la medida de las necesidades de sus clientes.

Entre los principales beneficios del uso del análisis del *Big Data* encontramos la reducción de los costes, por medio de la implementación de tecnologías de datos como Hadoop o las “*cloud-based analytics*” gracias a las cuales las empresas se ahorran costes relacionados con el almacenamiento; mejora en la toma de decisiones debido a la posibilidad de procesar y analizar información en tiempo real para tomar decisiones en vivo y para realizar predicciones del futuro; la creación de productos y servicios personalizados para cada cliente y la posibilidad de detectar más fácilmente posibles comportamientos sospechosos relacionados con el fraude.

Entre los principales riesgos o inconvenientes del *Big Data* encontramos los problemas relacionados con el almacenamiento, ya que para aprovechar las oportunidades que estas tecnologías nos pueden ofrecer una empresa debe tener una infraestructura de hardware y software adaptada; posibilidad de que los proyectos u aplicaciones acaben en fracaso, ya que acumular información no es una ventaja competitiva por sí misma, solo lo será si el análisis utilizado es el adecuado; problemas relacionados con la privacidad y la seguridad de los individuos por el uso indebido de la información que todos los usuarios y clientes comparten constantemente; y, finalmente, si los datos utilizados en el análisis no son los correctos esto puede repercutir de cara a los clientes y a la regulación vigente.

Como hemos mencionado hasta ahora, el objetivo de toda empresa y concretamente de las entidades bancarias, es recoger información en cantidades masivas. Es decir, buscan aprovechar sus fuentes de datos históricas y en tiempo real para poder analizarlas e integrarlas en sus sistemas operacionales y analíticos con el objetivo de obtener mejores directrices que permitan fomentar la creación de valor de su negocio.

El *Internet of Things* (IoT) se va a convertir en la principal fuente de información para todas las organizaciones y en la principal forma de conocer al cliente. Se trata de una tecnología en la que se integran sensores a dispositivos conectados a Internet que recogen información y luego la comparten por la red. Si los bancos son capaces de adaptarse y de implantar dispositivos capaces de identificar y conocer a un cliente a partir de datos biométricos, entonces verán mejoras en su productividad debido a un incremento de la eficiencia de las operaciones. Pero, además, también les permitirá acceder a nuevas ideas de negocio que redefinan el modelo bancario tradicional. Por otro lado, esta tecnología se también va a revolucionar la forma en la que las organizaciones gestionen sus relaciones

con los clientes. Los objetos que tengamos conectados a nosotros nos van a conocer, no solo nuestros gustos y nuestras preferencias, sino también nuestras rutinas, el contexto actual de cada persona que le lleve a realizar cierto tipo de acciones. Si las empresas son capaces de aprovechar este nuevo tipo de información, podrán no solo personalizar todavía más los productos para cada cliente, sino que, además, podrán jugar con otros factores como el precio y fijarlo en el momento adecuado al precio adecuado, siendo ésta la clave que garantizará su éxito.

El *Blockchain* ha aparecido como alternativa que permite a las empresas realizar y verificar transacciones financieras (como por ejemplo un pago), sin la necesidad de que haya una autoridad central que se involucre para garantizar su correcto funcionamiento. Esta tecnología ofrece la posibilidad de que las empresas refuercen sus estructuras y aceleren sus procesos y transacciones. Además, debido a que tiene una estructura descentralizada y segura, ofrece una aproximación distinta para que las organizaciones identifiquen los comportamientos fraudulentos. Finalmente, la Inteligencia Artificial es un sistema informático capaz de sentir, analizar y comprender la información y tomar medidas al respecto. Lo que se busca es crear un proceso capaz de realizar tareas que hasta entonces se pensaba que necesitaban de inteligencia humana. Se trata de una tendencia que está poco a poco implantándose en todos los sectores con visibles aplicaciones y en el caso del sector bancario permite cuantificar y ayudar a la gestión de riesgos.

La digitalización va a automatizar y a mejorar la gran mayoría de los procesos bancarios. Sin embargo, existen ciertos peligros subyacentes a estos cambios: nos encontramos en un mundo en el que las expectativas y las demandas de los clientes cada vez son mayores, y el avance tecnológico ha provocado y motivado la aparición de nuevos competidores significativos. La principal amenaza a la que se están enfrentando los bancos es la aparición de la digitalización y la falta de adaptación por parte de las entidades bancarias a ésta. La regulación y los cambios tecnológicos han permitido que nuevos jugadores y nuevas tecnologías entren en un mercado, donde el comportamiento del consumidor está en constante cambio, y cambien las reglas del juego.

Con lo cual, surge la pregunta ¿cómo será el banco del futuro? Considero que la única forma de que un banco tradicional asegure su supervivencia y éxito en el futuro es

mediante la flexibilidad. Si se vuelven lo suficientemente flexibles, los bancos serán capaces de adoptar una posición en la que pueda construir o externalizar servicios o productos y ser lo más competitivos posibles en todos los mercados. En el último apartado del trabajo hemos comentado las distintas tecnologías, respuestas y posibilidades que los bancos pueden y deben elegir para adaptarse al entorno cambiante. No obstante, no es suficiente con describir los procesos que pueden adoptar, es importante determinar los objetivos que pienso que deben seguir de cara al futuro, así como las futuras líneas de investigación:

1. Los Bancos, tienen que crear una red de canales diferenciada de manera a **fortalecer las relaciones con sus clientes**. La desaparición de los intermediarios conlleva a una etapa en la que los bancos tienen que segmentar a los clientes basándose en sus necesidades y crear proposiciones de valor para captar a la mayor cantidad de clientes.
2. Deben mejorar sus resultados mediante la reducción de costes, y la única forma de conseguirlo es promocionando el **cambio tecnológico** de toda su cadena productiva.
3. Deberán **redefinir sus productos y servicios**. Debido a la convergencia de los tipos de interés, los bancos deben redefinir su modelo de negocio y buscar ingresos por otros lados, asimismo proporcionando servicios de alto valor-añadido.
4. Tienen que buscar **transformarse constantemente**. Mejorar su gobierno, su infraestructura tecnológica, construir capacidades de análisis avanzadas, sus modelos de negocio, su estructura organizativa, su cultura, etc.
5. El factor **ético** como factor determinante. A pesar de que las empresas hagan firmar formularios de aceptación del uso de información confidencial, los clientes no los leen. Y debido a la importancia y a la trascendencia que está empezando a tomar, las empresas deberán ser capaces de adaptarse para evitar posibles riesgos regulatorios y legales.

6. Bibliografía

- Accenture. (2018). *Redefine Banking with Artificial Intelligence*. Accenture.
- Acens. (s.f.). *Bases de datos NoSQL. Qué son y tipos que nos podemos encontrar*. acenswhitepaper - Telefónica.
- Adam, K., Majid, M. A., & Zain, J. M. (Diciembre de 2014). *Big Data Management and Analysis*. Obtenido de ResearchGate: file:///C:/Users/Marta/Downloads/BigDataManagementandAnalysis.pdf
- Al Hujran, O., & Al-Debei, M. M. (2015). Big Data: Opportunities and Challenges. *The Fifth International Conference on Business Intelligence and Technology* (págs. 73 - 79). Amman: ResearchGate.
- Amakobe, M. (2015). *The impact of Big Data Analytics on the Banking Industry*. Colorado: Colorado Technical University.
- Aziz, S., & Dowling, M. (2018). *AI and machine learnig for risk management*. Rennes, Francia: SSRN Electronic Journal.
- Bănărescu, A. (2015). Detecting and Preventing Fraud with Data Analytics. *Procedia Economics and Finance*, 1827-1836.
- Balachandran, B. M., & Prasad, S. (2017). Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence. *International Conference on Knowledge Based and Intelligent Information and Engineering Systems* (págs. 1112-1122). Marsella : Elsevier.
- Barranco Fragoso, R. (18 de Junio de 2012). *¿Qué es Big Data? Todos formamos parte de ese gran crecimiento de datos*. Obtenido de IBM Developer: <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html>
- Bhadani, A. K., & Jothimani, D. (2016). *Big Data: Challenges, Opportunities and Realities*. Indian Institute of Technology Delhi, India.
- Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad*. (2019). Obtenido de PowerData: <https://www.powerdata.es/big-data>
- Datawarehouse*. (2019). Obtenido de Sinnexus Business Inteligence Informática Estratégica: https://www.sinnexus.com/business_intelligence/datawarehouse.aspx

- Deshpande, P., Siddiqi, A. H., Alam, K., & Parmar, K. (2016). Applications of Data Mining Techniques for Fraud Detection in Credit-Debit Card Transactions. *National Conference on Technological Advancement and Automatization in Engineering* (págs. 339-345). Mumbai: IJSRD.
- Duggal, P. S., & Paul, S. (December de 2013). Big Data Analysis: Challenges and Solutions. *International Conference on Cloud, Big Data and Trust* (págs. 269 - 276). Mesra, Ranchi, India: Research Gate. Obtenido de Research Gate.
- Ferris, A., Moore, D., Pohle, N., & Srivastava, P. (2014). Big Data: What is it, How is it collected and how might life insurance use it? *The Actuary Magazine*, 28-32.
- Fintech Network. (2019). *Four Blockchain Use Cases for Banks*. Fintech Network.
- Galimany Suriol, A. (2015). *La creación de valor en las empresas a través del Big Data*. Barcelona: Universidad de Barcelona.
- García Montalvo, J. (2014). El Impacto del Big Data en los servicios financieros. *Papeles de la Economía Española, Nuevos Negocios Bancarios*, 43-58.
- Goldman Sachs. (2014). *The Internet of Things: Making sense of the next mega-trend*. The Goldman Sachs Group, Inc.
- Gutierrez, D. D. (2018). *Big Data for Finance*. Obtenido de insidebigdata: <https://www.em360tech.com/wp-content/uploads/2017/10/Whitepaper-Big-Data-in-Finance.pdf>
- Hadi, H. J., Shnain, A. H., Hadishaheed, S., & Ahmad, A. H. (2015). Big Data and Five V's Characteristics. *International Journal of Advances in Electronics and Computer Science*, 16-23.
- Hadoop ¿Qué es y por qué es importante? (2019). Obtenido de SAS: https://www.sas.com/es_pe/insights/big-data/hadoop.html
- Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and Big Data Mining in Banking. *MDPI*, 1-13.
- Hernández Leal, E. J. (2016). *Aplicación de técnicas de análisis de datos y administración de Big Data ambientales*. Medellín, Colombia: Universidad Nacional de Colombia.

- Hernández-Leal, E. J., Duque-Méndez, N. D., & Moreno-Cadavid, J. (15 de Marzo de 2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *Technológicas*, pág. <http://www.scielo.org.co/pdf/teclo/v20n39/v20n39a02.pdf>.
- Herranz Gómez, R. (2014). *Bases de Datos NoSQL: Arquitectura y Ejemplos de Aplicación*. Madrid: Universidad Carlos III de Madrid.
- Historia Cronológica del Big Data*. (2019). Obtenido de WinShuttle: <https://www.winshuttle.es/big-data-historia-cronologica/>
- IBM. (Enero de 2016). *Banking on Blockchain: Charting the Progress of Distributed Ledger Technology in Financial Services*. Finextra Research. Obtenido de ING: <https://www.ingwb.com/media/1609652/banking-on-blockchain.pdf>
- IBM. (2019). *Bases de datos relacionales*. Obtenido de IBM: https://www.ibm.com/support/knowledgecenter/es/SSEPGG_8.2.0/com.ibm.db2.udb.doc/admin/c0004099.htm
- Kathuria, A. (2016). Impact of Big Data analytics on banking sector. *International Journal of Science, Engineering and Technology Research*, 3138-3141.
- Khodamoradi, M., & Mosa, R. N. (2016). The application of data mining techniques in risk management in banking industry. *Extensive Journal of Applied Sciences*, 163-167.
- Kubina, M., Varmus, M., & Kubinova, I. (2015). Use of Big Data for competitive advantage of company. *4th World Conference on Business, Economics and Management, WCBEM* (págs. 561 - 565). Elsevier.
- Latimore, D. (2018). *Artificial Intelligence in Banking*. Oliver Wyman.
- Liang, H., & Zhu, J. J. (Noviembre de 2017). *Big Data, Collection of (Social Media, Harvesting)*. Obtenido de ResearchGate: [file:///C:/Users/Marta/Downloads/LiangZhu2017Chapter%20\(1\).pdf](file:///C:/Users/Marta/Downloads/LiangZhu2017Chapter%20(1).pdf)
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big Data: The next frontier for innovation, competition and productivity*. McKinsey&Company.

- Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P., & Marrs, A. (2013). *Disruptive Technologies: Advances that will transform life, business and the global economy*. Mckinsey Global Institute - Mckinsey & Company.
- Marr, B. (25 de Febrero de 2015). *A brief history of big data everyone should read*. Obtenido de World Economic Forum: <https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read/>
- Martin, Á., Álvarez, M., Segovia, A. I., Casadas, V., Urbiola, P., Sánchez, A., . . . Sebastián, J. (2017). *The Internet of Things and Digital Banking*. BBVA.
- Maruti Techlabs. (2019). *Waret is Big Data Analytics and why do I need it for my business?* Obtenido de MARUTI TECHLABS: <https://www.marutitech.com/big-data-analytics-need-business/>
- Oguntimilehin, A., & Ademola, E. (2014). A Review of Big Data Management, Benefits and Challenges. *Journal of Emerging Trends in Computing and Information Sciences*, 433-438.
- Pascu, A. I. (2018). *Data Mining, Concepts and Applications in Banking Sector*. ACADEMICA BRÂNCUȘ.
- Press, G. (9 de Mayo de 2013). *A very Short History of Big Data*. Obtenido de Forbes: <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#6246298565a1>
- Pritzker, P. (September de 2015). *NIST Big Data Interoperability Framework: Volume 1, Definitions*. Obtenido de National Institute of Standards and Technology U.S. Department of Commerce: https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-1.pdf
- Pulakkazhy, S., & Balan, R. (2013). Data Mining in Banking and its applications - a review. *Journal of Computer Science*, 1252-1259.
- Raj S, S., & Jayanthi, S. (2015). *A Review of Data Mining Applications in Banking*. ResearchGate.

- Ramírez, A. (26 de Marzo de 2018). *¿Qué es la huella digital y por qué es importante conocerla?* Obtenido de InformaBTL: <https://www.informabtl.com/la-huella-digital-importante-conocerla/>
- Ravi, V., & Kamaruddin, S. (2017). Big Data Analytics Enabled Smart Financial Services: Opportunities and Challenges. *Springer International Publishing*, 15-39.
- SAS. (2019). *Big Data - What it is and why it matters*. Obtenido de SAS The Power to Know: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html#dmhistory
- Sauro, J. (4 de Marzo de 2014). *4 types of Customer Analytics Data to Collect*. Obtenido de Measuring U: <https://measuringu.com/customer-analytics/>
- Strohbach, M., Daubert, J., Ravkin, H., & Lischka, M. (2016). Big Data Storage. En J. D. Martin Strohbach, *New Horizons for a Data-Driven Economy* (págs. 119 - 141). Cavanillas.
- Subashini, B., & Chitra, K. (2013). Data Mining Techniques and its Applications in Banking Sector. *International Journal of Emerging Technology and Advanced Engineering*, 219-226.
- The five V's of big data*. (8 de May de 2017). Obtenido de BBVA: <https://www.bbva.com/en/five-vs-big-data/>
- Unión Internacional de Telecomunicaciones. (2012). *Serie Y: Infraestructura Mundial de la Información, aspectos del protocolo Internet y Redes de la Próxima Generación. Descripción General de Internet de los Objetos*. . Unión Internacional de Telecomunicaciones.
- Wei, W., Li, J., Cao, L., Ou, Y., & Chen, J. (25 de Junio de 2012). Effective detection of sophisticated online banking fraud on extremely imbalanced data. *Springer World Wide Web*. Obtenido de World Wide Web: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=63A4B753677EF59E71BCA6CE296D04B9?doi=10.1.1.705.2512&rep=rep1&type=pdf>
- What is Big Data?: The Three Vs*. (20 de July de 2016). Obtenido de Linux Academy Blog: <https://linuxacademy.com/blog/cloud/what-is-big-data-the-three-vs/>
- Wu, S. (2015). *Big Data Processing with Hadoop*. Turku University of Applied Sciences.

Yerpude, S., & Singhal, T. K. (2018). Internet of Things based Customer Relationship Management - A Research Perspective. *International Journal of Engineering and Technology*, 444-450.

Zarate Santovena, A. (June de 2013). *Big Data: Evolution, Components, Challenges and Opportunities*. Boston: Massachusetts Institute Of Technology. Obtenido de Massachusetts Institute of Technology: file:///C:/Users/Marta/Downloads/857767881-MIT.pdf